



HAL
open science

Orchestration et optimisation du cache dans les réseaux IoT

Sahand Khodaparas Talatapeh

► **To cite this version:**

Sahand Khodaparas Talatapeh. Orchestration et optimisation du cache dans les réseaux IoT. Autre [cs.OH]. Université d'Avignon; Université d'Urmia (Iran), 2024. Français. NNT : 2024AVIG0105 . tel-04953923

HAL Id: tel-04953923

<https://theses.hal.science/tel-04953923v1>

Submitted on 18 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**A dissertation submitted in partial fulfillment
for
the degree of Doctor of Philosophy**

Doctoral School: N° 536

- 1) Laboratoire Informatique d'Avignon (LIA)**
- 2) Electrical and Computer Engineering of Urmia University**

Specialty:

Computer Science and Computer Engineering

Presented by:

Sahand KHODAPARAS TALATAPEH

THESIS TITLE:

Cache Orchestration and Optimization in IoT Networks

Committee members:

Reviewers:

M. Antoine GALLAIS, INSA Hauts-de-France
Mme. Anna Maria VEGNI, Roma Tre University

Examiners:

M. Jamshid BAGHERZADEH, Urmia University
M. Yezekael HAYEL, Avignon University
Mme. Leila SHARIFI, Urmia University
M. Vahid SOLOUK, Urmia University of Technology

Advisors:

M. Abderrahim BENSLIMANE, Avignon University
M. Saleh YOUSEFI, Urmia University

Résumé

Dans le paysage en évolution rapide de l'Internet des objets (IoT) et de l'Internet des véhicules (IoV), la mise en cache émerge comme une technologie essentielle pour améliorer l'efficacité du réseau, réduire la latence et enrichir l'expérience utilisateur. Ces domaines technologiques sont confrontés à une demande croissante de mécanismes de gestion et de livraison des données plus performants, en raison de l'augmentation des volumes de données et de la complexité des réseaux.

Cette thèse explore des stratégies de mise en cache innovantes dans les domaines de l'IoT et de l'IoV, visant à améliorer les services réseau et les expériences utilisateur. La recherche présentée se divise en trois études distinctes mais interconnectées, chacune abordant des aspects critiques des performances du réseau, notamment la réduction de la latence, l'efficacité de la diffusion de contenu et l'expansion de la couverture réseau.

La première étude se concentre sur l'amélioration des capacités de mise en cache centrées sur le contenu dans les environnements IoT, en abordant les défis uniques posés par la nature hétérogène et dynamique de ces réseaux. Le paradigme de IoT implique la connexion et la communication de milliards de dispositifs diversifiés, ce qui complique la mise en œuvre de nouveaux services. Un aspect crucial du IoT est la nature centrée sur le contenu des données, où les applications se concentrent sur l'information elle-même plutôt que sur sa source ou son emplacement. Pour surmonter ces défis, nous adoptons une approche basée sur les réseaux de contenu (CCN), qui restructure l'architecture de l'Internet en nommant les données de manière à ce que chaque fragment devienne une unité auto-identifiable, facilitant ainsi la mobilité et l'accès par le cache.

Nous proposons une architecture hiérarchique de mise en cache utilisant un contrôleur SDN/Cache global (GSCC), qui centralise les décisions de mise en cache afin d'optimiser l'utilisation des ressources à travers le réseau. Les dispositifs IoT sont regroupés en clusters, avec des têtes de cluster agissant comme contrôleurs de cache locaux. Ces têtes de cluster rationalisent la diffusion du contenu en réduisant significativement les sauts de transmission nécessaires pour accéder aux données. La gestion des caches au sein de chaque cluster permet une distribution efficace du contenu en fonction des capacités énergétiques et de stockage des dispositifs, améliorant ainsi la durée de vie et la performance globale du réseau.

Pour déterminer les meilleures décisions de mise en cache, nous introduisons des techniques de prise de décision multicritères (MCDM) telles que le processus de hiérarchie analytique (AHP) et TOPSIS. Ces méthodes permettent d'évaluer diverses métriques liées aux opérations IoT, assurant une stratégie de mise en cache informée et adaptée à la nature dynamique du réseau. Par exemple, l'AHP permet de structurer les critères de décision en une hiérarchie, attribuant des poids relatifs à chaque critère basé sur leur importance, tandis que TOPSIS aide à identifier les meilleures options en comparant les alternatives à une solution idéale.

Les résultats de nos simulations démontrent l'efficacité de notre approche, avec un taux moyen de réussite du cache de 72 %, une réduction du nombre de sauts de 42 % et le maintien des fonctionnalités opérationnelles sur 60 % des nœuds pendant 90 % de la période de simulation. Ces résultats illustrent non seulement l'amélioration de l'efficacité de la mise en cache, mais aussi une réduction significative des délais d'accès aux données, répondant ainsi aux exigences critiques de latence et de disponibilité dans les environnements IoT.

En résumé, notre étude propose une solution innovante pour la gestion de la mise en cache dans les réseaux IoT, en intégrant des stratégies de décision multicritères et une orchestration hiérarchique des caches. Cette approche permet d'améliorer considérablement la performance et la durabilité des réseaux IoT, tout en répondant aux défis spécifiques posés par la diversité et la dynamique de ces environnements.

La deuxième étude examine les défis posés par la grande mobilité des véhicules dans les réseaux IoV. Pour répondre à ces défis, nous proposons deux méthodes de mise en cache spécialisées adaptées à des types de contenu distincts : la sécurité et l'infodivertissement. La méthode FM3C (Federated Learning-based Mobility-aware Collaborative Content Caching) est conçue pour l'infodivertissement, tandis que la méthode STAECC (Spatio-Temporal Characteristics Aware Emergency Content Caching) s'adresse aux données d'urgence. En utilisant l'apprentissage fédéré, nous pouvons prédire la popularité du contenu tout en préservant la confidentialité des utilisateurs. Ce processus est intégré à une prise de décision multicritères pour optimiser le placement du cache dans les unités routières (RSU).

La méthode STAECC utilise des techniques de prise de décision multicritères pour déterminer les contenus d'urgence à mettre en cache, en se basant sur leurs caractéristiques spatio-temporelles spécifiques. Les décisions sont prises localement par les RSU présentes sur le site de l'accident, ce qui permet de réduire les délais de traitement et d'améliorer la réactivité.

Pour le contenu d'infodivertissement, FM3C utilise des modèles de Long Short-Term Memory (LSTM) pour estimer la popularité du contenu selon l'historique des demandes des utilisateurs, tout en préservant la confidentialité grâce à l'apprentissage fédéré. Les véhicules forment des modèles locaux basés sur l'historique des demandes, et ces modèles sont ensuite agrégés par les RSU pour des décisions de cache efficaces. Cette approche garantit que le contenu le plus pertinent est mis en cache de manière proactive, améliorant ainsi l'expérience utilisateur en réduisant les temps de latence et en minimisant la charge sur le réseau principal.

Notre approche innovante améliore non seulement les taux de réussite du cache jusqu'à 98 %, mais réduit également les délais de récupération de contenu à moins de 10 ms, démontrant une fourniture efficace de contenu IoV. En outre, nous introduisons un serveur Fog pour coordonner les décisions de mise en cache entre les RSU, garantissant ainsi une gestion optimale des ressources réseau. Grâce à l'utilisation de réseaux définis par logiciel (SDN), nous pouvons transférer intelligemment les demandes vers les contenus mis en cache, permettant une coopération efficace entre les RSU et une réduction significative des délais de réponse.

En résumé, cette étude propose une solution robuste et scalable pour le cache de contenu dans les réseaux IoV, en utilisant des techniques avancées d'apprentissage fédéré et de prise

de décision multicritères pour répondre aux besoins spécifiques des contenus d'urgence et d'infodivertissement, tout en assurant une gestion efficace des ressources réseau et une amélioration de la qualité de service pour les utilisateurs.

L'étude finale présente une stratégie innovante de déploiement de drones (UAV) appelée SONA, qui prend en charge la mise en cache et étend la couverture réseau, notamment dans les scénarios IoV où les RSU sont débordées ou absentes. Utilisant la technologie blockchain pour l'orchestration décentralisée du réseau, SONA applique une approche d'optimisation de Pareto et transforme le défi de déploiement en un processus de décision de Markov (MDP), résolu à l'aide de l'apprentissage par renforcement profond (DRL). Cette stratégie réduit non seulement la latence moyenne à moins de 8 ms, mais permet également d'obtenir des taux de satisfaction des utilisateurs plus élevés avec moins de drones déployés, démontrant ainsi l'efficacité du positionnement dynamique des drones et de la mise en cache basée sur l'analyse prédictive.

La récente évolution des systèmes de transport intelligents (ITS) est marquée par l'avènement de IoV, inaugurant une nouvelle phase de transport intelligent et interconnecté. Cette transition nécessite des solutions avancées pour améliorer les services réseau et l'expérience utilisateur. Les technologies IoV ont catalysé le développement de nombreuses applications, allant de l'information sur le trafic à la planification des itinéraires et aux mécanismes de sécurité. En particulier, le domaine de l'infodivertissement en véhicule connaît une croissance substantielle. Avec l'arrivée des véhicules autonomes, les passagers peuvent profiter de contenus multimédia, des interactions sur les réseaux sociaux, du Metaverse, et de la réalité virtuelle. Les prévisions indiquent qu'en 2040, les véhicules autonomes représenteront environ 75 % du trafic routier, atteignant une valeur de marché potentielle de 7 000 milliards de dollars d'ici 2050.

Cependant, ces développements prometteurs présentent deux défis majeurs pour garantir une qualité de service élevée pour les utilisateurs IoV : la nécessité d'une faible latence et d'une couverture adéquate. Les applications IoV, notamment celles axées sur la sécurité, exigent des délais stricts, tandis que les applications de divertissement nécessitent à la fois une faible latence et une bande passante importante, mettant à rude épreuve le réseau de collecte si tout le contenu doit être récupéré depuis des serveurs distants. Le cache de contenu émerge comme une solution efficace, réduisant les transferts de données redondants et améliorant l'efficacité de transmission en plaçant le contenu populaire plus près des utilisateurs. De plus, l'infrastructure terrestre existante ne parvient souvent pas à couvrir entièrement les vastes réseaux routiers, une situation exacerbée par les catastrophes naturelles ou les volumes de demandes exceptionnellement élevés. Le déploiement de UAV apparaît comme une solution flexible pour renforcer l'infrastructure existante, étendant la couverture et fournissant une capacité supplémentaire là où elle est nécessaire.

La mise en cache efficace dans le cadre IoV repose sur l'observation que les utilisateurs d'une même zone géographique ont souvent des préférences de contenu similaires. En cachant stratégiquement ce contenu populaire plus près des utilisateurs, il est possible de réduire considérablement les transferts de données répétitifs et de raccourcir les délais de réponse. Cependant, la mise en œuvre d'une stratégie de cache efficace n'est pas sans défis, notamment la sélection du contenu à mettre en cache en raison des ressources limitées. Une approche prédictive s'avère nécessaire pour estimer la popularité future du contenu, en se

basant sur les modèles historiques de demandes des utilisateurs, tout en respectant leur vie privée.

Pour surmonter ces défis, notre méthodologie utilise l'apprentissage fédéré, modifié pour améliorer la sélection des participants et l'agrégation des modèles. Cette approche permet de rassembler des informations sans accéder directement aux données sensibles des utilisateurs, assurant ainsi un mécanisme de prédiction de la popularité des contenus respectueux de la vie privée. En parallèle, nous optimisons le déploiement des UAVs pour répondre aux limitations de l'infrastructure terrestre.

SONA optimise à la fois le nombre et les emplacements des UAVs en utilisant des modèles de LSTM pour prédire les schémas de trafic des véhicules. En transformant ce défi en un MDP et en utilisant le DRL, nous élaborons une stratégie de déploiement optimale qui maximise la connectivité et le débit tout en minimisant le nombre de drones nécessaires. Cette approche améliore non seulement la couverture et la qualité de service, mais réduit également la consommation d'énergie.

Enfin, notre stratégie utilise la technologie blockchain pour créer un cadre décentralisé et transparent permettant la coopération entre RSU et UAV. Cette coopération optimise les stratégies de cache et de déploiement des UAV, assurant une distribution efficace et diversifiée des contenus mis en cache, tout en maximisant la couverture et la qualité de service. La technologie blockchain facilite un mécanisme de consensus collectif, garantissant que les stratégies de déploiement des UAVs reflètent les besoins actuels du réseau.

En résumé, cette recherche propose une solution avancée et scalable pour le cache de contenu et le déploiement des UAVs dans les réseaux IoV, utilisant des techniques de prédiction avancées, d'optimisation et de blockchain pour améliorer la qualité de service et l'efficacité du réseau.

Ensemble, ces études illustrent une progression de la mise en cache et de la gestion de réseau centralisées vers des systèmes de réseau plus autonomes, décentralisés et intelligents. Les recherches futures pourraient explorer l'amélioration des mesures de sécurité de ces mécanismes pour se protéger contre l'évolution des cybermenaces et le développement de modèles d'incitation pour encourager la participation des utilisateurs, en particulier dans les scénarios d'apprentissage fédéré. En abordant ces domaines, nous pouvons améliorer davantage la robustesse et l'efficacité des écosystèmes IoT et IoV, en garantissant qu'ils puissent répondre aux demandes croissantes des infrastructures de transport et de communication modernes.

Mots clés: Caching, IoT, IoV, UAV, Federated learning, Blockchain

Abstract

In the rapidly evolving landscape of the Internet of Things (IoT) and the Internet of Vehicles (IoV), caching emerges as a pivotal technology to enhance network efficiency, reduce latency, and improve user experiences. These technological domains face growing demands for better data management and delivery mechanisms due to increasing data volumes and network complexity.

In this thesis, we explore innovative caching strategies within the realms of the IoT and the IoV to enhance network services and user experiences. The research presented spans three distinct yet interconnected studies, each addressing critical aspects of network performance, including latency reduction, content delivery efficiency, and network coverage expansion.

The first study focuses on enhancing content-centric networking caching capabilities within IoT environments. By employing hierarchical network orchestrations and a global SDN/Cache controller (GSCC), our approach centralizes cache decisions, optimizing resource usage across the network. Devices are grouped into clusters with heads acting as cache controllers, streamlining content delivery and reducing transmission hops significantly. We introduce Multi-Criteria Decision Making (MCDM) techniques such as Analytical Hierarchy Process (AHP) and TOPSIS, which evaluate multiple metrics to optimize caching decisions. Simulation results from this study show an average cache hit rate of 72%, a reduction in hop counts by 42%, and the maintenance of operational functionality across 60% of nodes during 90% of the simulation period.

The second study delves into the challenges posed by the high mobility of vehicles in IoV networks. We propose two specialized caching methods tailored to distinct content types prevalent in IoV: safety and infotainment. The Federated Learning-based Mobility-aware Collaborative Content Caching (FM3C) method is designed for infotainment, while the Spatio-Temporal Characteristics Aware Emergency Content Caching (STAECC) method caters to emergency data. We utilize federated learning to predict content popularity while preserving user privacy, integrated with multi-criteria decision-making to optimize cache placement in Road-Side Units (RSUs). Our innovative approach not only enhances cache hit rates to up to 98% but also significantly reduces content retrieval delays to under 10ms, showcasing an efficient provision of IoV content.

The final study introduces an innovative UAV deployment strategy termed SONA, which supports caching and extends network coverage, particularly in IoV scenarios where RSUs are overwhelmed or absent. Utilizing blockchain technology for decentralized network orchestration, SONA applies a Pareto Optimization approach and transforms the deployment challenge into a Markov Decision Process (MDP), solved using Deep Reinforcement Learning (DRL). This strategy not only reduces average latency to less than 8ms but also achieves higher user satisfaction rates with fewer UAVs deployed, demonstrating the effectiveness of dynamic UAV positioning and caching based on predictive analytics.

Together, these studies illustrate a progression from centralized caching and network management towards more autonomous, decentralized, and intelligent network systems. Future research could explore enhancing the security measures of these mechanisms to protect against evolving cyber threats and developing incentive models to encourage user participation, particularly in federated learning scenarios. By addressing these areas, we can further advance the robustness and efficiency of IoT and IoV ecosystems, ensuring that they can meet the growing demands of modern transportation and communication infrastructures.

Keywords: Caching, IoT, IoV, UAV, Federated learning, Blockchain

Remerciements

I am profoundly grateful to everyone who has supported me throughout the journey of completing this thesis. The guidance, encouragement, and insight provided have been essential in navigating the complexities of my research. Their unwavering support has been a pillar of strength during this challenging but rewarding process.

First of all, I would like to express my deep gratitude to my thesis director, Mr. Abderahim BENSLIMANE, Professor at the University of Avignon, and my supervisor, Mr. Saleh YOUSEFI, Professor at the University of Urmia, for their support, their precious advice, and their scientific contributions throughout these years of my thesis.

I would also like to express my gratitude and deep respect to Prof. Anna Maria VEGNI and Prof. Antoine GALLAIS, who accepted to judge this work and to be the reporters. I extend my thanks to all the members of the jury for the great attention they have given to my work.

Special thanks are due to my wife, whose unconditional love and encouragement have been fundamental to my success. Her patience and understanding have made this achievement as much hers as it is mine. Thank you for being my constant source of inspiration and for standing by me through every step of this journey. Your support has truly been the cornerstone of my progress.

Copyright license

Cache Orchestration and Optimization in IoT Networks © 2024 by Sahand Khodaparastalatapeh is licensed under Creative Commons Attribution 4.0 International

Contents

Résumé	ii
Abstract	vi
Remerciements	viii
Copyright license	viii
1. Introduction	1
1.1 Context and Motivation	1
1.2 Research Objectives and Contributions	2
1.2.1. Objective 1: Hierarchical Caching Scheme for IoT Devices	2
1.2.2. Objective 2: Caching in IoV with a Focus on Reducing Delay	2
1.2.3. Objective 3: Integration of IoV Caching Schemes with UAVs and Blockchain Technology	3
1.3 Thesis Structure	3
1.4 List of publications	5
2. Background and Literature Review	6
2.1 Internet of Things Paradigm	6
2.1.1. IoT Applications	6
2.1.2. IoT Challenges	7
2.2 The Internet of Vehicles (IoV): Bridging Connectivity on the Move	8
2.2.1. Exploring IoV Architectures	9
2.2.2. IoV Applications: Enhancing Safety, Comfort, and Traffic Management	11
2.2.3. Challenges in Realizing the Potential of IoV	13
2.3 The Evolution and Versatility of Unmanned Aerial Vehicles in Wireless Communication Networks	14
2.3.1. Enhancing Connectivity through UAVs: A Dual Approach to Wireless Communication Networks	15
2.3.2. Identifying Key Challenges in IoV Networks	16
2.4 Exploring the Paradigm Shift with Blockchain Technology	17
2.4.1. Key Characteristics of Blockchain Technology	18
2.4.2. Enhancing IoT Through Blockchain Integration	19
2.5 The Evolution and Impact of Caching in Network Technologies	20
2.5.1. Caching in IoT: Foundations and Evolution	21
2.5.2. Advancements in Caching for IoV Systems	27
2.5.3. Integrating UAVs with Caching Strategies: A New Horizon	33
3. A software-defined caching scheme for the Internet of Things	40
3.1 Introduction	40
3.2 Research Motivation and Contribution	41
3.3 System Model	42
3.3.1. Challenges	42
3.3.2. Proposed Model	43
3.4 Inter-cluster global decision making	45

3.5 Intra-cluster local decision machining	49
3.5.1. Node selection	50
3.5.2. Determining candidate contents	54
3.6 Performance evaluation	55
3.7 Conclusion	61
4. An Intelligent Caching Scheme Considering the Spatio-Temporal Characteristics of Data in Internet of Vehicles	62
4.1 Introduction	62
4.2 System Model	64
4.2.1. Mobility Model	66
4.2.2. Content Retrieval Mechanism	68
4.3 Proposed Caching Scheme	68
4.3.1. Emergency Content	69
4.3.2. Infotainment Content	70
4.4 Popularity Prediction, Based on Federated Learning	71
4.4.1. Participant Selection	73
4.4.2. Global Model Download	74
4.4.3. Local training	74
4.4.4. Model Aggregation	75
4.5 Multi Criteria Decision Making	76
4.6 Global Content Placement	77
4.7 Performance Evaluation	77
4.7.1. Infotainment Content Caching	77
4.7.2. Emergency Content Caching	82
4.8 Conclusion	85
5. Enhancing IoV with SONA: A Synergistic Approach to UAV Deployment and Blockchain-Based Caching	86
5.1 Introduction	86
5.2 System Model	89
5.2.1. Caching Model	90
5.2.2. User Satisfaction Model	91
5.2.3. Traffic Model	91
5.2.4. UAV Mobility Model	92
5.2.5. UAV Energy Consumption Model	92
5.2.6. Channel Model	93
5.2.7. Transmission Model	95
5.2.8. Blockchain Model	96
5.3 Proposed Caching Algorithm	97
5.3.1. FM3C	97
5.3.2. Blockchain-Enabled Caching Strategy	98
5.4 UAV Deployment Strategy	99
5.4.1. Traffic and Data Rate Estimation Methodology	99
5.4.2. Formulation of the Problem	100
5.4.3. A Reinforcement Learning Approach to the Solution	102
5.5 Performance Evaluation	104
5.5.1. Simulation Environment Description	104
5.5.2. Simulation Parameters	105

5.5.3. Simulation Results	106
5.6 Conclusion	112
6. Conclusion and Future Directions	113
Bibliography	116

1. Introduction

1.1 Context and Motivation

The Internet of Things (IoT) [1] represents a paradigm shift in the digital transformation of our world, where the interconnectivity of devices extends the internet beyond the traditional confines of computing devices, to a broader range of everyday objects. This interconnectedness enables objects to collect and exchange data, facilitating a level of communication and automation previously deemed unattainable [2]. IoT finds application across various domains, including but not limited to smart homes [3], healthcare [4], agriculture [5], and transportation [6], significantly enhancing efficiency, sustainability, and convenience in everyday activities. Among its various applications, the Internet of Vehicles (IoV) [7] emerges as a notable subcategory, epitomizing IoT in mobility contexts. The IoV extends the IoT's connectivity and data exchange capabilities to vehicles, enabling dynamic information sharing among vehicles, infrastructure, and pedestrians to improve traffic management, enhance safety, and personalize the driving experience [8]. This distinctiveness positions the IoV not only as a pivotal IoT application but also as a vanguard in the mobility-driven digital landscape.

Despite the transformative potential of IoT and IoV, their implementation surfaces critical challenges in data management and network efficiency. The proliferation of IoT devices and the resultant data explosion impose a considerable strain on backhaul networks [9], necessitating innovative solutions to manage the surge in content generation and requests. This challenge is further compounded by the dispersed nature of IoT devices, which often necessitates ubiquitous network coverage to ensure seamless connectivity [10]. While these challenges are somewhat universal across IoT applications, specific subcategories face unique obstacles. For instance, conventional IoT devices, predominantly sensors, grapple with resource constraints, including limited battery life and intermittent connectivity due to their basic communication capabilities [11]. Ensuring data availability, even when the source IoT node is momentarily inaccessible, becomes paramount in such scenarios.

Conversely, the IoV domain is particularly challenged by the need for reduced latency to support applications critical to safety and the complexities introduced by the mobility of vehicles [12], which can result in inconsistent connections. Furthermore, the deployment of interactive applications within the IoV ecosystem accentuates the need for innovative solutions tailored to this dynamic environment.

Caching emerges as a potent solution to these challenges [13], proposing a method to locally store frequently accessed data to alleviate the load on backhaul networks and ensure data availability despite potential connectivity issues. However, the effective implementation of caching within the IoT and IoV landscapes is fraught with challenges. Deciding which contents to cache and which to omit requires a nuanced understanding of the network's dynamics and the varying nature of IoT subcategories. This complexity underscores the necessity for bespoke caching solutions that are meticulously tailored to each IoT subcategory, taking into account their unique challenges and objectives.

In response to these nuanced challenges, this thesis develops distinct caching methodologies, each optimized for specific IoT and IoV environments. Initially, we focus on traditional IoT devices, proposing caching strategies that mitigate their inherent limitations. Subsequently, we delve into the IoV realm, devising specialized caching approaches catered

to the diverse application types within this domain. Recognizing that caching efficiencies are moot without robust connectivity, we explore the deployment of Unmanned Aerial Vehicles (UAVs) as a means to bolster network coverage and enhance caching capabilities across these varied IoT landscapes. This multifaceted approach not only addresses the prevailing challenges in data management and network efficiency but also paves the way for a more integrated, efficient, and responsive IoT and IoV ecosystem.

1.2 Research Objectives and Contributions

This thesis aims to address the critical challenges in data management and network efficiency within the IoT and the IoV, as identified in the preceding motivation section. Central to this endeavor is the development of efficient caching schemes tailored to the unique requirements of each scenario, thereby enhancing system performance, data availability, and user experience. Our research is guided by the following specific objectives:

1.2.1. Objective 1: Hierarchical Caching Scheme for IoT Devices

Our first objective involves proposing a caching scheme optimized for IoT devices characterized by limited capabilities and reliance on battery power. The overarching goals of this initiative are to minimize the load on the backhaul network, fulfill the majority of requests locally, extend the lifespan of IoT nodes, and improve data availability. To achieve these aims, we adopted a hierarchical approach to caching decisions. This methodology employs distinct criteria at each layer for selecting suitable IoT nodes for content caching and determining the contents to be cached. Given the diversity of criteria, multiple-criteria decision-making (MCDM) [14] techniques serve as the foundation for our decisions, ensuring a balanced and efficient caching strategy.

Contributions from Objective 1:

- Development and implementation of a hierarchical caching decision framework for IoT devices.
- Application of MCDM techniques to optimize content selection and node designation, enhancing system efficiency and node longevity.

1.2.2. Objective 2: Caching in IoV with a Focus on Reducing Delay

Our second objective examines caching within the IoV, targeting the reduction of delays to foster the emergence of new applications and functionalities. Recognizing the dichotomy of IoV content into safety and infotainment categories, we proposed two distinct caching methods. The first method is specifically designed for safety-related content, emphasizing the spatio-temporal characteristics of such data. The second method focuses on infotainment content, utilizing popularity prediction to identify optimal content for caching. A key aspect of this work is the emphasis on privacy preservation during the popularity prediction process, ensuring user data confidentiality.

Contributions from Objective 2:

- Introduction of two specialized caching schemes for IoV, addressing the unique needs of safety and infotainment content.
- Integration of privacy-preserving mechanisms in popularity prediction, safeguarding user data while optimizing caching decisions.

1.2.3. Objective 3: Integration of IoV Caching Schemes with UAVs and Blockchain Technology

The final objective explores the integration of the proposed IoV caching schemes with UAVs and blockchain technology. This initiative aims to extend caching capacities and coverage through the strategic deployment of UAVs, while leveraging blockchain to decentralize and secure caching decisions. A UAV deployment strategy is developed to maximize the Quality of Service (QoS) for users and minimize the associated costs by deploying an optimized number of UAVs, ensuring an effective and cost-efficient expansion of network capabilities.

Contributions from Objective 3:

- Deployment of a novel UAV-based strategy to enhance network coverage and caching capacity within the IoV.
- Implementation of blockchain technology to facilitate decentralized, secure caching decisions, ensuring robustness and transparency.

Through these objectives and the corresponding contributions, this thesis presents a comprehensive approach to overcoming the challenges of caching and network management in IoT and IoV environments. By tailoring caching schemes to specific device capabilities and content types, and innovatively integrating emerging technologies like UAVs and blockchain, this work significantly advances the state-of-the-art in IoT and IoV systems.

1.3 Thesis Structure

This thesis is meticulously structured to address the complexities and challenges of implementing efficient caching strategies within the IoT and the IoV landscapes. Following the introduction, the thesis is organized into five additional chapters, each focusing on distinct but interconnected areas of research. Below is a summary of the content and primary focus of each chapter, linked to the research objectives and contributions outlined previously:

Chapter 2: Background and Literature Review

This chapter lays the foundational groundwork for the thesis by providing an extensive review of the literature, establishing the necessary background to understand the current state of research in Internet of Things (IoT) and Internet of Vehicles (IoV) caching schemes. Initially, we explore various concepts such as IoT, IoV, Unmanned Aerial Vehicles (UAVs), and blockchain technology, discussing their applications and the challenges associated with them. Following this, the focus shifts to caching mechanisms, which are divided into three main subsections: general IoT caching schemes, IoV-specific caching schemes, and studies exploring the role of UAVs in caching. Each subsection begins with a thorough introduction of related

works. After presenting these studies, we identify the existing research gaps and categorize the approaches taken by various researchers. This discussion sets the stage for the novel contributions of this thesis by highlighting areas in need of further exploration and innovation.

Chapter 3: A Software-Defined Caching Scheme for the Internet of Things

Aligned with the first objective, this chapter delves into the development of a caching scheme for IoT devices with limited capabilities and reliance on battery, aimed at minimizing backhaul network load, responding to most requests locally, and extending node lifetimes. It details the hierarchical method for caching decision-making, where different criteria are employed at each layer for selecting suitable IoT nodes and contents for caching. Utilizing MCDM techniques like the AHP and TOPSIS, the chapter outlines how these approaches underpin our caching decisions, significantly enhancing system efficiency and data availability.

Chapter 4: An Intelligent Caching Scheme Considering the Spatio-Temporal Characteristics of Data in the Internet of Vehicles

Focusing on the second objective, this chapter explores caching within the IoV, aiming to reduce delays and facilitate the introduction of new applications. It presents two distinct caching methods tailored for safety and infotainment contents, addressing their unique characteristics. The FM3C method targets infotainment content, while the STA ECC method is devised for emergency content. This chapter discusses the use of LSTM models for popularity prediction and the implementation of privacy-preserving mechanisms, ultimately demonstrating the effectiveness of these methods in improving cache hit rates, reducing delays, and enhancing the Quality of Experience (QoE) for users.

Chapter 5: Enhancing IoV with SONA: A Synergistic Approach to UAV Deployment and Blockchain-Based Caching

Addressing the final objective, this chapter presents the integration of the proposed IoV caching schemes with UAVs and blockchain technology. It highlights how this novel approach aims to extend caching capacities and coverage by deploying UAVs and leveraging blockchain for decentralized and secure caching decisions. The chapter elaborates on the mathematical optimization model and machine learning algorithms employed for dynamic UAV deployment and caching strategies, showcasing significant network efficiency improvements through adaptive positioning of UAVs and content caching based on predictive analytics. The use of blockchain technology is emphasized for enhancing the IoV ecosystem's robustness, facilitating autonomous interaction among established infrastructures and UAVs for optimized decision-making.

Chapter 6: Summary, Concluding Remarks, and Future Research Directions

The final chapter summarizes the key findings and contributions of the thesis, reflecting on the implications of the research for advancing IoT and IoV network services. It provides concluding remarks that encapsulate the essence of the work and discusses potential avenues for future research, highlighting opportunities for further advancements in caching strategies and network efficiency within IoT and IoV environments.

1.4 List of publications

This thesis is based on the following publications submitted in international journals:

1. **S. Khodaparas**, A. Benslimane, and S. Yousefi, "A software-defined caching scheme for the Internet of Things," *Computer Communications*, vol. 158, pp. 178–188, May 2020, doi: [10.1016/j.comcom.2020.05.002](https://doi.org/10.1016/j.comcom.2020.05.002). [15]
2. **S. Khodaparas**, A. Benslimane, and S. Yousefi, "An Intelligent Caching Scheme Considering the Spatio-Temporal Characteristics of Data in Internet of Vehicles," *IEEE Trans. Veh. Technol.*, pp. 1–15, 2023, doi: [10.1109/TVT.2023.3337051](https://doi.org/10.1109/TVT.2023.3337051). [16]
3. (Submitted to IEEE Internet of Things Journal). **S. Khodaparas**, A. Benslimane, and S. Yousefi, "Enhancing IoV with SONA: A Synergistic Approach to UAV Deployment and Blockchain-Based Caching,"

Other publications by the author, presented at International Conferences:

1. **S. Khodaparas**, S. Yousefi, and A. Benslimane, "A Multi Criteria Cooperative Caching Scheme for Internet of Things," in ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China: IEEE, May 2019, pp. 1–6. doi: 10.1109/ICC.2019.8761546. [17]

2. Background and Literature Review

2.1 Internet of Things Paradigm

The concept of the IoT was first introduced in 1999 by Kevin Ashton. He envisioned IoT as a system bridging the gap between the physical world and the Internet, utilizing RFID technology and pervasive sensor devices to observe, identify, and interact with real-world entities [18]. This initial definition has since evolved, with numerous interpretations being offered [19]-[22]. However, one of the most comprehensive definitions describes IoT as "An open and comprehensive network of intelligent objects that have the capacity to auto-organize, share information, data, and resources, acting and reacting in the face of situations and changes in the environment." [23] This definition underscores the inherent capacity of IoT devices for intelligence, interconnectivity, and adaptability, marking a significant leap towards autonomous, smart environments.

The widespread adoption of IoT devices and their integration into the fabric of daily life represent a significant technological evolution. Recent data, including Cisco's annual report [24], indicates that machine-to-machine (M2M) connections account for 50% of all global connections, a trend that is expected to continue its upward trajectory. This underscores the increasing reliance on and the importance of IoT technologies in facilitating automated and seamless communication across various sectors and applications.

Moreover, projections by HS Market Telecommunication Industry [25] suggest a more ambitious future, estimating that by the year 2030, approximately 125 billion IoT devices will be connected to networks worldwide. This exponential increase in IoT devices competing for network resources presents not only a considerable challenge to the already scarce network resources but also introduces new dimensions to QoS and data management. These developments highlight the critical need for innovative solutions to accommodate the rapid growth of IoT devices while ensuring efficient, reliable, and secure data communication and management.

2.1.1. IoT Applications

IoT technology has found applications across various domains, each tailored to leverage the unique capabilities of interconnected devices to address specific challenges and enhance efficiency, safety, and quality of life. Some prominent use cases include:

- **Smart Healthcare:** Utilizing IoT devices for monitoring patient health in real-time, facilitating remote diagnostics, and improving patient care through data-driven insights and interventions [26]-[28].
- **Smart Environment:** Leveraging sensor networks within the IoT framework transforms the approach to environmental monitoring, enabling a more proactive stance toward environmental protection and management. Beyond measuring air quality, water quality, and soil conditions, IoT technologies play a pivotal role in disaster detection and wildlife conservation. These systems are crucial for the early detection of natural disasters, including forest fires, earthquakes, tsunamis, and avalanches, allowing for timely evacuation and response efforts to mitigate damage and save lives. Additionally, IoT devices facilitate the tracking of wild animals, contributing to research on their migration patterns, health, and behavior, thereby enhancing efforts in wildlife conservation and the

protection of endangered species [29]-[34].

- **Smart Cities:** Integrating IoT technologies to optimize urban services such as lighting, traffic management, and waste collection, enhancing urban living through improved efficiency, reduced costs, and increased sustainability [35]-[37].
- **Smart Transportation:** The advent of IoT technologies in the realm of Intelligent Transportation Systems (ITS) has revolutionized the way traffic is managed, significantly reducing congestion and enhancing road safety. At the heart of this transformation is the IoV, a cornerstone application that brings the power of IoT to vehicles and transportation infrastructures. It fosters a network of dynamic, real-time communication between vehicles, road infrastructures, and traffic control centers, streamlining traffic management and safety protocols. Beyond the critical improvements in traffic flow and safety, the IoT ecosystem significantly elevates the in-vehicle experience through the integration of sophisticated infotainment systems. These systems offer passengers access to a myriad of services and content, including live traffic updates, weather forecasts, multimedia streaming, and interactive navigation aids, all tailored to make journeys more enjoyable and informative. By harnessing the capabilities of IoT for in-vehicle infotainment, smart transportation not only prioritizes efficiency and safety but also places a high value on enhancing the user experience, thereby redefining the future of travel. [38]-[41].

Each of these use cases exemplifies the potential of IoT to revolutionize industries by harnessing the power of data and connectivity. As IoT continues to evolve, its applications are set to expand, further integrating into various facets of human activity and significantly impacting how we live, work, and interact with our environment.

2.1.2. IoT Challenges

As the IoT continues to evolve and expand, it confronts a spectrum of challenges that could impede its growth and the realization of its full potential. Addressing these challenges is critical for ensuring the seamless integration and functionality of IoT systems across various domains. Below, we delve into some of the key challenges that the IoT faces:

- **Battery Life:** A significant portion of IoT devices relies on battery power, with some even powered by renewable energy sources that can be unpredictable. Given the distributed nature of these devices, frequently replacing or recharging batteries is impractical. Thus, enhancing battery life and developing energy-efficient devices that can operate for extended periods become paramount. Innovative energy harvesting technologies and power management techniques are being explored to prolong device operability without human intervention [42][43].
- **Interoperability:** The essence of IoT lies in its ability to not only connect devices to the Internet but also facilitate meaningful M2M communication. Achieving this level of interoperability among a diverse array of devices, each possibly using different protocols and standards, poses a considerable challenge. The development and adoption of globally accepted standards and protocols are crucial for enabling seamless communication and interaction across the IoT ecosystem, thereby unlocking the potential for advanced integrated applications [44]-[46].
- **Coverage:** IoT devices are dispersed across a wide range of environments, from dense indoor spaces to expansive outdoor areas. Ensuring consistent and extended coverage to maintain reliable connectivity for all these devices, irrespective of their

location, poses a significant challenge. The issue becomes particularly pronounced in remote or rural areas, where conventional network infrastructure may be sparse or entirely absent. To overcome these challenges, advancements in network technologies, such as low-power wide-area networks (LPWANs), have been instrumental. LPWANs offer broader coverage and deeper penetration capabilities at reduced costs [47]-[49]. Additionally, the deployment of UAVs is emerging as a promising solution to extend coverage dynamically [50]-[52]. UAVs can be rapidly deployed to areas lacking infrastructure, acting as temporary network nodes that bridge coverage gaps. This flexibility makes UAVs an invaluable asset in ensuring ubiquitous connectivity for IoT devices across varied and challenging terrains, further expanding the potential reach and utility of IoT applications.

- **Big Data:** The proliferation of IoT devices is set to generate an immense volume of data, with a recent report indicating that data output will reach 79.4 zettabytes (ZB) by 2025 [63]. This vast quantity of data presents significant challenges in terms of storage, processing, and analysis [53]. Managing this big data effectively requires robust data analytics tools and technologies capable of extracting valuable insights in real-time [54]. Furthermore, ensuring the privacy and security of this data amidst its vastness adds another layer of complexity to its management. A promising solution to these challenges is the adoption of edge computing [55]-[57]. By processing data at the edge of the network, closer to where it is generated, and responding to requests locally, edge computing can significantly reduce the need for data to travel back and forth between distant servers and IoT devices. This not only alleviates the strain on network resources but also enhances data processing speed and efficiency. Additionally, edge computing can contribute to improved data privacy and security by limiting the exposure of sensitive information and reducing the potential attack surface for cyber threats. Through these mechanisms, edge computing offers a viable approach to mitigating the challenges associated with big data in the IoT ecosystem.
- **Security and Privacy:** As the number of connected devices grows, so does the surface for potential cyberattacks. Ensuring the security of IoT devices and the privacy of the data they handle is of utmost importance. Challenges arise from the heterogeneity of devices, many of which may lack the necessary computational power to implement standard security protocols. Developing lightweight, robust security solutions that can be deployed across various devices is critical for protecting against breaches and ensuring user privacy [58][59].
- **Routing:** Efficient data transmission within the IoT network, especially in scenarios involving constrained devices and networks, is crucial. Routing in IoT involves managing the data flow from numerous devices over networks that may be highly dynamic and resource-constrained. Designing adaptive routing protocols that can efficiently handle the scalability, variability, and bandwidth constraints of IoT networks is essential for maintaining optimal network performance and reliability [60]-[62].

2.2 The Internet of Vehicles (IoV): Bridging Connectivity on the Move

The concept of Vehicular Ad-hoc NETWORKS (VANET) lays the foundational stone for

understanding the more evolved and comprehensive framework of the IoV. VANET is established as a network formed ad-hoc among moving vehicles and other connected entities through a wireless medium, facilitating the exchange of critical information [64]. This network supports Vehicle-to-Vehicle (V2V) [65], Vehicle-to-Infrastructure (V2I) [66], and Vehicle-to-Road (V2R) communications [67], heralding a new era in ITS. Vehicular Ad-hoc Networks (VANETs) unlock a myriad of applications ranging from blind crossing assistance [68] and safety enhancement [69] to traffic condition monitoring and collision prevention [70][71].

Recent statistics illuminate the burgeoning scale of vehicular networks. In 2018 alone, vehicle production reached an impressive 92 million units [72], and according to the International Organization of Motor Vehicle Manufacturers, the global count of vehicles in use has soared to approximately 1.59 billion [73]. This vast and ever-increasing number of vehicles underscores the critical need for advanced vehicular communication and management systems.

Evolving from the integration of VANET and IoT principles, the IoV emerges as a pivotal concept aimed at enhancing road safety, reducing accident rates, and streamlining traffic management. Vehicles within the IoV paradigm are not merely transit tools but are transformed into intelligent objects. These objects are equipped with sophisticated software and hardware capable of performing complex calculations and real-time decision-making based on the analysis of incoming data. Furthermore, IoV extends the connectivity of these smart vehicles beyond traditional vehicular networks. It incorporates heterogeneous network access [74], encompassing a broad spectrum of users, vehicles, things, and networks, thereby facilitating universal internet access and enabling a wide range of services from traffic management to infotainment.

The IoV represents a significant leap towards creating smarter, safer, and more efficient transportation systems. By leveraging the combined capabilities of VANETs and the IoT, the IoV offers a robust platform for vehicles to interact with each other and with the surrounding infrastructure. This interaction paves the way for a future where vehicles can autonomously navigate, communicate potential hazards, optimize traffic flows, and provide passengers with seamless connectivity and entertainment options, making the journey as important as the destination.

2.2.1. Exploring IoV Architectures

The evolution of vehicular networks into IoV has been met with significant interest from academia and industry, leading to efforts aimed at standardizing architectural frameworks to support this advanced ecosystem. Among these, the European Telecommunications Standards Institute (ETSI) has been at the forefront, proposing architectures that adapt and extend beyond traditional models like TCP/IP or OSI to meet the specific requirements of VANETs and IoV systems [75].

A distinguishing feature of the ETSI architecture, as illustrated in Figure 2.1, is the inclusion of a facilities layer. This layer is pivotal for supporting VANET-related applications such as Cooperative Awareness Messages (CAM), Decentralized Notification Messages (DENM), and the Local Dynamic Map (LDM), enhancing both communication processes and cooperative vehicular functionalities. Further deviations from conventional models include the amalgamation of network and transport layers into a single cohesive layer and the introduction of dedicated layers for management and security, underscoring the emphasis on

robust, secure, and efficient network operations within the IoV ecosystem.

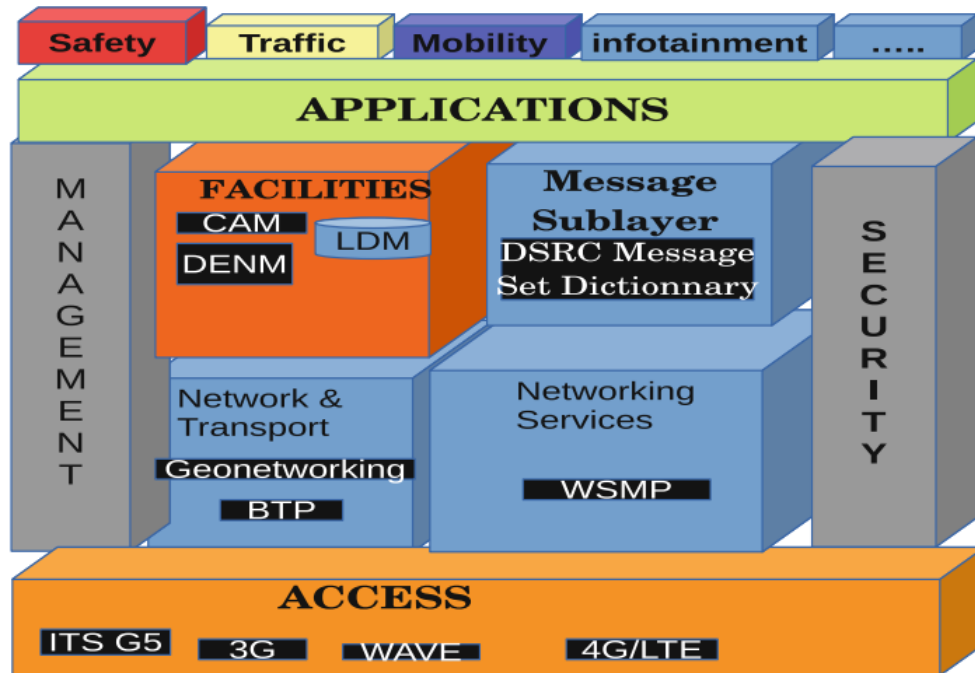


Figure 2.1. ETSI architecture.

Moreover, the architecture incorporates an ITS-specific stack that integrates GeoNetworking addressing, offering a tailored approach to vehicular communication that prioritizes geographic information and routing. However, while these architectural elements provide a solid foundation for internal device functionalities and communication, there exists a gap in addressing the dynamics of external device interactions, such as handover processes and other mechanisms that necessitate seamless interoperability between different devices and networks.

The integration of IoT technologies into VANETs has unlocked a plethora of promising solutions and applications, from real-time autonomous driving applications to comprehensive road traffic management and enhanced comfort applications. This convergence has given rise to the IoV, a sophisticated ecosystem characterized by four main components:

- **End Points:** This includes vehicles, smartphones, sensors, and other connected devices, forming the network's operational nodes.
- **Infrastructure:** Comprising Roadside Units (RSUs), Wi-Fi hotspots, and cellular networks (5G/LTE) base stations, the infrastructure facilitates the core connectivity and communication framework.
- **Operations:** Encompassing policy enforcement, flow-based management, security, and pricing strategies to ensure the network's optimal performance and integrity.
- **Services:** Including a variety of cloud services (public, private, and enterprise) that provide subscription-based or enterprise-level data, voice, or video services.

In addressing the architectural evolution towards facilitating IoV deployment, Bonomi from Cisco [76] has proposed a four-layer architecture that delineates the essential stages of IoV

communication: Embedded Systems and Sensors, Multi-service Edge, Core, Data Center, and Cloud. This framework, as depicted in Figure 2.2, outlines a structured approach to processing and managing the complex data flows and interactions inherent in IoV systems, from the initial data collection at the edge to the comprehensive processing and storage in data centers and cloud services.

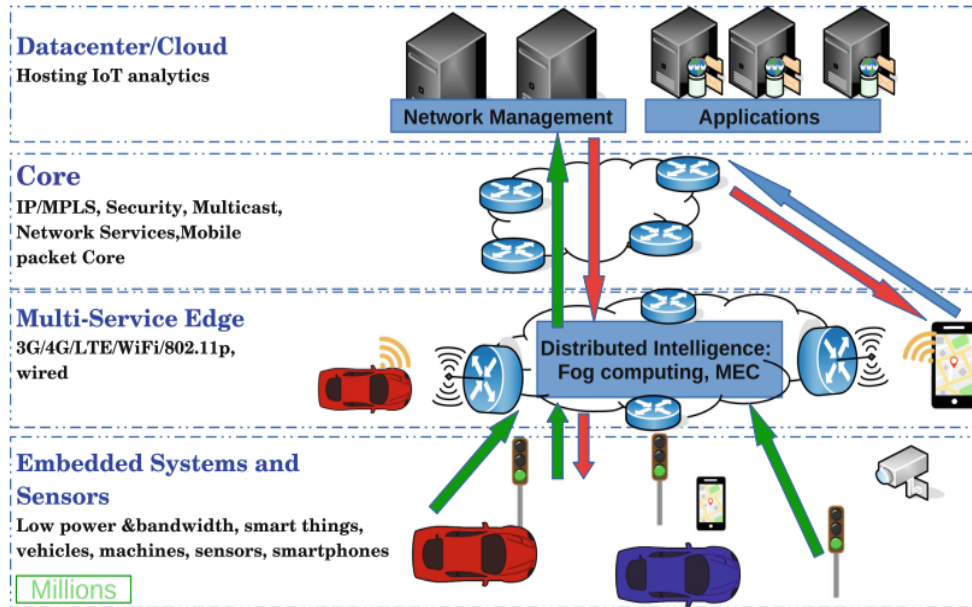


Figure 2.2. IoV 4 layers architecture based on [76].

The continued exploration and refinement of IoV architectures by industry and academia are crucial for unlocking the full potential of vehicular networks. By fostering a deeper understanding of these frameworks and addressing the challenges of external device interactions, the IoV can achieve greater interoperability, security, and efficiency, driving forward the advancement of intelligent transportation systems.

2.2.2. IoV Applications: Enhancing Safety, Comfort, and Traffic Management

The realm of the IoV encompasses a broad spectrum of applications aimed at transforming the driving experience, enhancing road safety, and optimizing traffic management. These applications can be categorized into three primary domains, each with its specific objectives and functionalities:

1) Safety-Related Applications

At the forefront of IoV applications are those designed to enhance road safety. The core aim of safety-related applications in IoV is to mitigate road accidents, a concern that carries the weight of potential loss of life. These applications are characterized by their low tolerance for delays, as they are intended to act preemptively, offering drivers advance warnings to prevent mishaps. Examples of safety-related applications include:

- **Traffic Signal Violation Warning:** Alerts drivers about potential red light and stop sign violations [77].

- **Left Turn Assistant:** Provides guidance for safe left turns at intersections [78].
- **Intersection Collision Warning:** Warns of possible collisions at intersections [79].
- **Pedestrian Crossing Information:** Informs drivers about pedestrians crossing or about to cross the road [80].
- **Emergency Vehicle Signal Preemption:** Helps clear the path for emergency vehicles by controlling traffic signals [81].
- **Post-Crash Warning:** Notifies approaching vehicles of recent accidents to prevent secondary collisions [82].
- **Wrong Way Driver Warning:** Alerts drivers and traffic management centers about vehicles moving in the wrong direction [83].
- **Blind Spot Warning Position:** Provides warnings about vehicles or obstacles in the driver's blind spot [84].
- **Cooperative Adaptive Cruise Control:** Adjusts the vehicle's speed to maintain safe distances from other vehicles [85].
- **Road Condition Warning:** Informs drivers about hazardous road conditions like ice, potholes, or debris [86].

2) Comfort and Infotainment Applications

Designed to enhance the comfort and entertainment of travelers, infotainment applications within IoV can accommodate some level of message loss and delays. These applications enrich the travel experience by providing access to multimedia content, internet services, and location-based services. Examples include:

- **Intelligent Parking Navigation System:** Assists drivers in finding parking spots and navigating parking areas [87].
- **Internet Service Provisioning:** Offers on-the-go internet access for various online services [88].

Additional subcategories cover a range of services from finding nearby amenities like ice cream parlors, movie theaters, and coffee shops to enabling online video streaming, carpooling, and file downloading.

3) Traffic Efficiency and Management

Applications under this category aim to streamline traffic flow and reduce congestion, thereby preventing accidents and saving travel time. By informing vehicles about current and upcoming traffic conditions, these applications allow for route adjustments to avoid congested areas. Key applications include:

- **Intersection Management:** Optimizes intersection flow to reduce wait times and improve efficiency [89].
- **Road Congestion Management:** Provides real-time traffic condition updates to help drivers avoid congested routes [90].
- **Digital Map Downloading:** Allows for the downloading of updated digital maps for accurate navigation and routing [91].

These diverse IoV applications collectively work towards creating a safer, more enjoyable, and efficient driving environment. By leveraging advanced communication technologies and data analytics, IoV extends the capabilities of traditional vehicular networks, paving the way for innovative solutions to longstanding challenges in transportation.

2.2.3. Challenges in Realizing the Potential of IoV

The IoV presents a revolutionary approach to enhancing vehicular communication and operations. However, the realization of its full potential is not without challenges. These obstacles span various aspects of network operations, data management, and service delivery, each contributing to the complexities of deploying and managing IoV systems effectively.

1) Network Scalability and Connectivity

One of the fundamental challenges facing IoV is ensuring scalable and reliable connectivity amidst the ever-increasing number of connected vehicles. The expansion of IoV networks, expected to encompass millions of vehicles globally, demands a network infrastructure that can handle massive data traffic while maintaining seamless communication among vehicles, infrastructure, and other entities [92].

2) Data Volume and Management

IoV generates a significant volume of data through its diverse applications, from safety alerts to traffic updates and beyond. Efficiently managing this influx of data—ensuring its storage, processing, and analysis—is crucial for the operational efficiency of IoV systems. The big data challenge in IoV underscores the need for sophisticated data management solutions that can handle the scale and complexity of the data involved [93].

3) Latency and Real-Time Processing

For IoV applications, particularly those related to safety and traffic management, low latency and real-time processing are imperative. The effectiveness of these applications is heavily reliant on the timely transmission and processing of data, with any delay potentially compromising the system's safety and efficiency [94].

4) Reliability and Consistency

Ensuring data reliability and consistency across IoV's dynamic environment is another significant challenge. With vehicles constantly on the move and network conditions changing, providing consistent access to accurate and current information is essential for the effective operation of IoV applications [95].

5) Network Coverage and Accessibility

Achieving comprehensive network coverage and accessibility, especially in areas with limited infrastructure, presents another hurdle for IoV. Ensuring that vehicles can access necessary data and services regardless of their location is critical for the universality and effectiveness of IoV systems [96].

6) Infotainment Content Delivery

Beyond operational challenges, delivering infotainment content within IoV introduces its own set of issues. The high volume of content and its interactive nature demand substantial network resources, potentially leading to congestion and degraded service quality. Ensuring efficient delivery of infotainment content, therefore, becomes a challenge that needs addressing within the IoV framework [97].

Together, these challenges highlight the complexities involved in implementing and managing IoV systems. Addressing these issues is crucial for unlocking the transformative potential of IoV, paving the way for a future where vehicular networks enhance safety, efficiency, and the overall driving experience.

2.3 The Evolution and Versatility of Unmanned Aerial Vehicles in Wireless Communication Networks

UAVs have emerged as a revolutionary technology, reshaping the landscape of wireless communication networks. These aerial devices offer a promising solution to establish rapid and reliable connectivity, especially in areas where traditional terrestrial communication infrastructures fall short. UAVs, operating as mobile aerial base stations, extend emergency communication services [98] to ground terminals in various critical scenarios including battlefields, disaster zones, congested roads, blind spots, and rural regions [99]. The agility and mobility of UAVs allow for swift deployment, closely serving users and adapting to fluctuating demand, which stands as a significant advantage over fixed terrestrial base stations.

The utilization of UAVs in enhancing wireless coverage is multifaceted. From serving as flying cell sites that expand the reach of static territorial base stations to acting as aerial base stations for geographical areas beyond cellular networks, UAVs are at the forefront of innovation in network technology. Notable initiatives like Verizon's airborne LTE service [100] and Project Loon's balloon-based flying base stations [101] highlight the diverse applications of UAVs in providing critical connectivity solutions. This adaptability to varying traffic loads without the constraints of permanent deployment locations underpins the growing interest among wireless carriers in using UAVs for opportunistic coverage enhancement during large-scale events.

Advancements in technology have significantly contributed to the evolution of UAVs, making them lighter, more flexible, and capable of longer operational periods. These improvements have expanded UAV applications beyond military and public safety to encompass a wide range of civil uses. The integration of UAVs into 5G and beyond networks underscores their potential to boost network capacity through quick deployment, low maintenance costs, and unmatched mobility. Furthermore, the deployment of UAVs into existing wireless communication networks offers a seamless solution without necessitating cell reorganization, providing a practical approach to enhancing network services.

Military applications of UAVs have evolved from border surveillance and reconnaissance to include public services by agencies such as police and public safety departments [102], aiding in disaster response and environmental monitoring. The versatility of UAV sizes allows for their use in solo missions or in formations, demonstrating their potential in civilian applications. This diversification is supported by advancements in electronics and sensor technology, broadening the scope of UAV network applications to traffic monitoring, wind estimation, and remote sensing [103].

The broader adoption of UAVs extends to various civilian and commercial applications, leveraging their reduced costs and sizes for weather monitoring, aerial imaging, surveillance, cargo transportation, and disaster recovery efforts [104]-[106]. The reliability and security of wireless communication with controllers are critical for the efficient and safe navigation of UAVs. As the landscape of wireless communications evolves, UAV applications continue to expand, including data collection in wireless sensors and IoT networks [107], wireless power transfer [108], mobile edge computing [109], and enhancing terrestrial network communications with intelligent reflecting surfaces [110].

The evolution and versatility of UAVs in wireless communication networks underscore their significant role in overcoming connectivity challenges and enhancing network capacity. Their rapid deployment, adaptability, and broad range of applications from military to civilian uses

mark UAVs as a pivotal technology in the future of wireless communications.

2.3.1. Enhancing Connectivity through UAVs: A Dual Approach to Wireless Communication Networks

1) Expanding Horizons: UAVs in Infrastructure-less Environments.

In scenarios lacking pre-existing terrestrial infrastructure, such as remote or rural areas or in the aftermath of disasters where ground base stations (BSs) are non-functional as shown in Fig. 2.3, UAVs step in as a critical technology. These UAVs provide essential wireless coverage directly to users or act as aerial relays to facilitate communication between otherwise obstructed nodes.

Use Cases and Benefits:

- **Maximizing Radio Coverage:** UAVs can significantly extend the reach of wireless networks into areas that lack infrastructure, ensuring that more users gain access to communication services [111].
- **Enhancing Quality of Service (QoS):** Through strategic deployment, UAVs can ensure a higher portion of users meet QoS constraints, improving user experience [112].
- **Rate Maximization:** The aim of rate maximization when utilizing UAVs in wireless networks is twofold: ensuring all users have access to a basic level of connectivity and maintaining a steady, reliable service quality despite network fluctuations. This approach not only guarantees equitable resource distribution but also stabilizes the user experience, making it both fair and consistently satisfactory [113].
- **Network Performance Optimization:** Enhancing network performance through UAV deployment focuses on maximizing the system's sum-rate through spectral efficiency, which optimizes the use of the available spectrum. Concurrently, efforts to maximize the System Signal-to-Interference Ratio (SIR) and minimize the decoding error probability/rate aim to bolster the reliability of communications across the network. Additionally, reducing transmission latency is a key objective, particularly vital for supporting applications where time is of the essence [114].

2) Enhancing Network Efficacy: UAVs in Concert with Ground Infrastructure

In this scenario, UAVs complement and enhance existing cellular networks. They assist in managing network load, especially in congested areas or where direct communication is hindered by physical obstructions. UAVs also facilitate backhaul connectivity for small cells.

Use Cases and Benefits:

- **System Sum-Rate Maximization:** By offloading traffic from ground BSs, UAVs help increase the overall data throughput of the network [115].
- **System Spectral Efficiency Maximization:** Optimizing the use of available spectral resources, UAVs contribute to a more efficient and capable network [116].
- **System Capacity and Coverage Probability Maximization:** Integrating UAVs into

the network infrastructure can significantly increase the network's capacity while also improving the likelihood of coverage across all areas [117].

- **Individual User's Rate Maximization:** Targeted deployment can improve service for users with specific needs, ensuring optimal data rates on a per-user basis [118].
- **Minimum and Average SIR Maximization:** By providing direct line-of-sight connections, UAVs can enhance the signal quality, reducing interference and improving overall network performance [119].

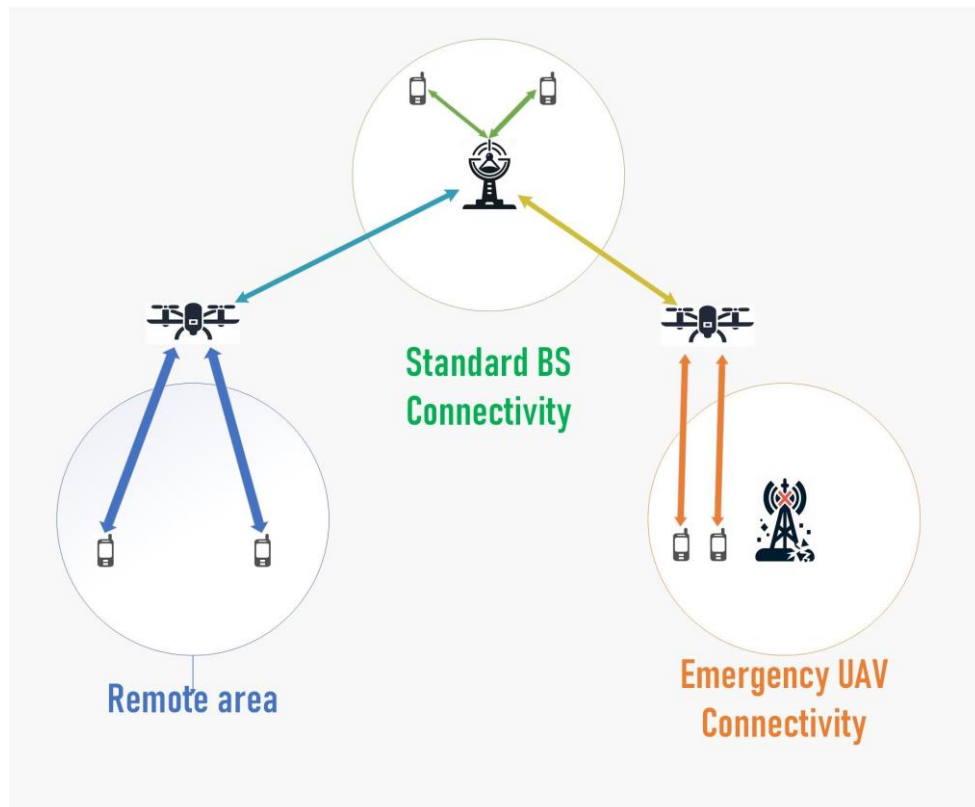


Figure 2.3. UAVs deployment in Infrastructure-less Environments

2.3.2. Challenges in UAV-Enabled Communication Networks

The rapid evolution of UAV technology, fueled by advancements in aerodynamics, electronics, and energy efficiency, has significantly expanded their capabilities and applications. However, integrating UAVs into wireless communication networks presents unique challenges that must be addressed to fully leverage their potential.

- **Operation Time Limit**
A significant constraint for UAVs is their limited battery life, which directly impacts their operational duration and, by extension, the continuity of services they provide. Extending the operational time of UAVs while ensuring they perform their intended functions effectively is a key challenge in network design [120].
- **Operational Costs**
The integration of UAVs into communication networks entails considerable operational costs, from deployment and setup to maintenance and management.

Balancing these financial investments with the aim of delivering quality services is a challenge that impacts the scalability and sustainability of UAV-enabled networks [121].

- **Interference and Network Stability**

Interference from various sources, including natural obstacles, other electronic devices, and competing network signals, can significantly impact communication quality and network stability, necessitating a flexible and scalable approach to network architecture [122].

- **Meeting Diverse User Demands**

UAV networks are envisioned to support a wide array of services, each with distinct requirements regarding latency, bandwidth, and reliability. Catering to these diverse and sometimes conflicting demands poses a significant challenge, necessitating a flexible and scalable approach to network architecture [123].

- **Security and Privacy**

The deployment of UAVs raises substantial security and privacy concerns, as they often handle sensitive data and operate in environments susceptible to cyber threats. Establishing robust security measures and protocols is imperative to protect against unauthorized access and ensure the privacy of communicated information [124].

Addressing these challenges is fundamental to the successful integration of UAVs into communication networks, paving the way for their broader application across various domains. Solutions that tackle these issues will not only enhance the capabilities of UAV-enabled networks but also expand their potential to revolutionize the way we connect and communicate.

2.4 Exploring the Paradigm Shift with Blockchain Technology

Blockchain technology represents a foundational shift in how data is stored, verified, and transferred across the digital landscape. At its core, blockchain functions as a distributed database solution or a decentralized public ledger system that chronologically records data records or transactions. This technology's inherent attributes—decentralization, transparency, immutability, and auditability—have not only revolutionized the domain of cryptocurrencies but have also extended its influence across various sectors.

Foundational Concepts and Evolution

Originally conceived to support Bitcoin, the first cryptocurrency introduced by Nakamoto in 2009 [125], blockchain has evolved far beyond its initial application. It operates on a chain of blocks, each storing transactions recorded on a public ledger, growing continuously as new blocks are appended. This decentralized framework is powered by core technologies like digital signatures [126], cryptographic hash functions [127], and distributed consensus algorithms [128], facilitating peer-to-peer transfers of digital assets without the need for intermediaries. The decentralized nature of blockchain eliminates the need for third-party validation, enhancing the system's security and transparency. All nodes within the network share access to the transaction history, ensuring information trustworthiness and immutability.

Broadening Horizons: Beyond Cryptocurrencies

Blockchain's impact extends across finance, healthcare, government, manufacturing, and distribution, innovating and transforming applications ranging from supply chain management and digital media transfers to remote service delivery and identity management [129]. Its application in voting systems, data storage, and renting [130] showcases the potential for blockchain to secure and streamline various governmental and business processes. Furthermore, blockchain facilitates distributed resources management, crowdfunding, electronic voting, and the governance of public records [131]-[133], underscoring its versatility and capacity to redefine existing operational paradigms.

The Advent of Smart Contracts

A significant advancement within the blockchain ecosystem is the development of smart contracts. Conceptualized by Nick Szabo in 1993 [134] as computerized transaction protocols executing contract terms, smart contracts have found their ideal platform in blockchain technology. These contracts self-execute or enforce contractual clauses, introducing efficiency, precision, cost reduction, and transparency to transactions. The emergence of Blockchain 2.0, characterized by integrated smart contract functionality, has further expanded blockchain's applicability. Smart contracts are stored on the blockchain, identifiable by unique addresses, and operate through transactions sent to these addresses. Their execution is guaranteed by the blockchain consensus protocol, fostering the development of innovative applications in diverse areas.

As blockchain technology continues to evolve, its capacity to serve as the backbone for a wide array of applications highlights its significance in the digital era. The convergence of blockchain with smart contracts and its application across industries signify a transformative shift, promising a future where decentralized, transparent, and secure transactions form the basis of digital interactions.

2.4.1. Key Characteristics of Blockchain Technology

Blockchain technology, by design, embodies a set of inherent characteristics that distinguish it from traditional centralized systems and enable its revolutionary applications across various sectors. These key attributes form the bedrock of blockchain's security, efficiency, and reliability.

- **Decentralization**

Unlike traditional centralized transaction systems, where transactions are validated through a central authority (e.g., a central bank), blockchain operates on a peer-to-peer (P2P) basis. This decentralization allows transactions to be conducted directly between two parties without the need for authentication by a central agency. By circumventing centralized servers, blockchain significantly reduces server costs, including both development and operational expenses, and alleviates performance bottlenecks typically associated with central points of control.

- **Persistency**

Blockchain's architecture ensures that once a transaction is confirmed and recorded in a distributed ledger, altering it becomes nearly impossible. Each block in the chain contains a hash of the previous block, creating an immutable sequence of

transaction records. This linkage makes any attempt to tamper with transaction data easily detectable by other nodes in the network. The process of modifying information would require changing the hash data in all preceding blocks, a task so computationally intensive that it is deemed impractical. Consequently, blockchain's persistency provides an immutable and tamper-proof ledger, enhancing the trustworthiness and security of data recorded on the network.

- **Anonymity**
Blockchain allows users to interact within the network using addresses generated for each transaction. This feature enables users to maintain privacy and avoid identity exposure, as there is no central entity storing personal information. However, it's important to note that the level of anonymity can vary, particularly in private blockchains [135], where transaction visibility may be restricted to certain network participants, altering the degree of privacy afforded by the system.
- **Auditability**
The distributed ledger of a blockchain network meticulously records all transactions, complete with digital timestamps. This comprehensive record-keeping facilitates the auditability and traceability of transactions. In networks like Bitcoin, transactions can be traced iteratively, providing transparency and facilitating the audit process. However, techniques such as tumbling [136], which involves moving money through numerous accounts, can complicate the tracing of funds back to their origin, presenting challenges in linking transactions to their initial source.

Together, these characteristics define the essence of blockchain technology, offering a decentralized, secure, and transparent framework for digital transactions. By leveraging these attributes, blockchain has the potential to revolutionize not only financial transactions but also a wide array of applications in various industries, setting the stage for a new era of digital interaction and exchange.

2.4.2. Enhancing IoT Through Blockchain Integration

The convergence of IoT and blockchain technology heralds a transformative era in digital communication and infrastructure. Blockchain, with its inherent characteristics of decentralization, immutability, and transparency, offers a robust solution for many of the challenges faced by the IoT ecosystem. This integration fosters a trusted environment where information is not only reliable but also securely shared among numerous participants, enhancing both security and traceability across various applications.

- **Revolutionizing Traceability and Security**
One of the paramount benefits of integrating blockchain with IoT is the enhancement of data traceability and security. In sectors like food safety [137], where ensuring the integrity of product information is crucial, blockchain can facilitate exhaustive traceability involving numerous stakeholders, from production to distribution. Similarly, in smart cities and smart cars [138], sharing reliable data securely can significantly improve service quality and foster the adoption of new technologies, making blockchain an indispensable ally for IoT.
- **Decentralization and Scalability**
The shift from centralized systems to P2P distributed architectures, enabled by blockchain, eliminates central points of failure and alleviates bottlenecks. This decentralization not only enhances fault tolerance and system scalability but also

mitigates the risk of data monopolization by a few entities. By reducing IoT silos, blockchain contributes to the overarching goal of improving IoT scalability, facilitating a more interconnected and efficient digital ecosystem.

- **Autonomy in IoT**

Blockchain empowers the creation of smart autonomous assets, paving the way for the development of hardware as a service and device-agnostic applications. This autonomy allows IoT devices to interact directly with each other without server intermediation, enhancing application flexibility and device interoperability. Such advancements are crucial for the evolution of IoT into a more adaptive and responsive network.

- **Reliability of IoT Information**

The immutable nature of blockchain ensures that IoT data remains secure and unaltered over time. Participants in the network can verify data authenticity with confidence, knowing that the information has not been compromised. This level of reliability is critical for applications requiring sensor data traceability and accountability, where the integrity of data is paramount.

- **Security Enhancements**

Blockchain can significantly bolster the security of IoT communications by treating device message exchanges as validated transactions through smart contracts. This approach secures communications between devices, optimizing current security protocols used in IoT. By incorporating blockchain, IoT networks can achieve enhanced security, ensuring that communications and data exchanges are protected against unauthorized access and tampering.

The integration of IoT and blockchain represents a critical leap forward in building a more secure, transparent, and efficient digital world. By addressing key challenges such as scalability, autonomy, reliability, and security, this convergence has the potential to redefine the landscape of digital communication and infrastructure, ushering in a new era of interconnected and intelligent systems.

2.5 The Evolution and Impact of Caching in Network Technologies

Caching, a concept that significantly enhances the efficiency of data retrieval processes, has its roots deeply embedded in the history of computing and has evolved remarkably over the decades. From its early implementations in the 1960s computer systems, such as the innovative virtual memory system of the Atlas Supervisor for the Atlas Computer [139], caching mechanisms have been pivotal in optimizing the use of main memory and storage. These initial steps laid the foundational principles of caching, which would undergo significant transformations to address the evolving needs of computing and network technologies.

The adoption of caching extended beyond the confines of computing systems into the realm of the Internet, marking a pivotal shift in how content was distributed and accessed. Instead of relying solely on central servers for content retrieval, popular webpages began to be replicated in smaller, strategically placed servers (caches) around the globe. This development aimed at achieving three critical objectives: reducing network bandwidth usage, decreasing content access time, and alleviating server congestion. The explosion of Internet traffic in the late 1990s further complicated cache management, catalyzing the emergence of

Content Delivery Networks (CDNs). As an integral component of today's Internet ecosystem, CDNs utilize advanced monitoring and control techniques to manage an interconnected network of caches, ensuring efficient content delivery across vast geographical distances.

In contemporary network and communication technologies, caching mechanisms are ubiquitous, permeating various domains and architectures. From enhancing storage in future Internet architectures and information-centric networks to enabling caching in 5G wireless systems, caching has demonstrated its capacity to benefit both network infrastructure by reducing operational costs and end-users by improving service quality. Recently, the focus has shifted towards leveraging caching to ameliorate content delivery within wireless networks. As the physical layer access rate enhancement and the deployment of additional base stations become increasingly costly, these traditional methods struggle to keep pace with the rapidly escalating mobile data traffic. Caching emerges as a promising solution to bridge this gap, offering a cost-effective approach to boosting network capacity and enhancing user experiences.

As we delve deeper into the specifics of caching within IoT, IoV, and UAV-assisted networks in the following subsections, it becomes evident that caching is not just a supplementary technology but a cornerstone of modern network infrastructure. Its evolution from a rudimentary memory management technique to a sophisticated tool for optimizing network performance and content delivery underscores the dynamic nature of caching as an essential element in the progression of network technologies.

2.5.1. Caching in IoT: Foundations and Evolution

This subsection delves into the crucial role of caching within the IoT ecosystem. It aims to explore how caching techniques have been foundational in addressing the inherent challenges of IoT, such as limited power in IoT devices and the excessive load on backhaul networks due to the vast amount of data produced. By reviewing related work, we will examine how caching strategies have evolved to mitigate these issues, enhancing the efficiency of IoT systems despite the inherent limitations of device power and network capacity.

Study [140] adopts an advanced approach to in-network caching within the IoT, capitalizing on the capabilities of Content-Centric Networking (CCN) to address the unique challenges of IoT environments. The authors propose a Multi-attribute In-Network Caching Decision (MACD) algorithm that considers a comprehensive set of attributes for making caching decisions, including data freshness, node energy level, available storage size, and hop count. This innovative approach aims to optimize resource utilization in constrained IoT devices, ensuring data is cached where it will be most effective and accessible. However, the methodology introduces certain limitations, such as the lack of cooperation between nodes, leading to potential redundancy in cached content. Furthermore, the algorithm does not account for the network topology or the nodes' relative locations, which could enhance caching efficiency by considering the physical proximity of nodes to requesters. Additionally, the focus on intermediary nodes for caching decisions may prevent requests from benefiting from content cached on non-intermediary nodes, potentially missing opportunities for optimizing data retrieval times and network resource use.

The work presented in [141] introduces a topology-aware caching mechanism for Information-Centric Networking (ICN) in the IoT, leveraging Approximate Betweenness Centrality (ABC) to make caching decisions without extensive network topology knowledge or high communication overhead. This approach aims to optimize limited IoT device resources by strategically caching content at nodes with high betweenness centrality, thus reducing the need for nodes to access content from distant sources and saving energy. However, this

strategy overlooks the redundancy of cached content across the network, assumes request and response paths are identical (which may not always be true) and treats all content equally without considering content-specific metrics like popularity or size.

Reference [142] unveils an innovative caching strategy within the IoT domain, capitalizing on Named Data Networking (NDN) to selectively cache data based on its popularity and the remaining lifetime. This approach smartly navigates the challenges posed by the ephemeral nature of IoT data, ensuring that only the most demanded and currently relevant information is stored at the network's edge. By enabling NDN nodes to autonomously determine the caching worthiness of data, the strategy adeptly reduces latency and conserves network resources. However, this method doesn't directly tackle the issue of cache redundancy, potentially leading to suboptimal cache space utilization. Moreover, it overlooks the impact of network topology and node capabilities on caching efficiency.

Research in [143] optimizes edge caching for the IoT through a novel application of Deep Reinforcement Learning (DRL), circumventing the need for preconceived knowledge on data popularity or user request patterns. This approach acknowledges the transient nature of IoT data and incorporates communication costs, calculated based on bandwidth consumption (affected by content size) and retrieval delays, to refine caching decisions dynamically. Despite its innovative strides, the method overlooks the potential for smart routing to enhance the utility of cached data further. Moreover, the feasibility of deploying DRL on IoT nodes, given their constrained computational resources, remains a pertinent concern. Additionally, the adaptability of this model to the volatile popularity trends characteristic of IoT content usage was not thoroughly examined, underscoring crucial avenues for future investigation.

Reference [144] proposes an innovative caching strategy for IoT data, leveraging NDN to optimize data retrieval across both edge and core network segments. The study introduces two distinct caching strategies: a coordinated approach for edge domains and an autonomous strategy for core segments. Despite both strategies considering content popularity and freshness as key metrics, their implementation diverges in operational mechanics. In the core network, caching decisions are made autonomously by each node, aiming to reduce unnecessary data forwarding and improve network efficiency. Conversely, at the edge, a novel mechanism is introduced whereby a **CACHED** flag is set to true by the first node that caches the content, aiming to reduce redundancy by preventing subsequent edge nodes from caching the same content. However, the approach exhibits several shortcomings. It simplifies the network hierarchy into just two levels, without considering potentially valuable metrics such as node centrality, which could further refine caching decisions based on network topology. Additionally, while it successfully mitigates redundancy in the edge network, it does so at the possible cost of optimal cache placement; the first node to cache a piece of content, often being closer to the source, may not be the most strategic location for minimizing latency for future requests. Furthermore, the absence of smart routing mechanisms to direct requests towards nodes where content is already cached might limit the potential benefits of reduced latency and network traffic, particularly in scenarios where cached content is distributed unevenly across the network.

In the paper [145], the authors introduce an innovative approach to caching in the IoT environment by deploying a discrete multi-agent variant of the Soft Actor-Critic (SAC) algorithm. This methodology allows multiple edge nodes (ENs) to autonomously make caching decisions, guided by the goal of minimizing the long-term average weighted cost that encompasses Age of Information (AoI), transmission energy consumption, and network traffic demands. Central to their strategy is a policy orchestrated by the cloud, which aims to reduce the overall network costs by making efficient caching decisions. When a data request is

initiated, the nearest EN first checks its local cache. If the content is either outdated or absent, the EN evaluates whether to retrieve the latest data from the cloud or directly from IoT sensors, as determined by the SAC algorithm, ensuring data freshness and network resource optimization. However, the proposed method exhibits certain limitations, such as a lack of explicit cooperation among ENs, which could lead to disparities in storage and load distribution. Moreover, the algorithm does not account for the varying capabilities of nodes or the characteristics of the content, such as size, focusing primarily on content freshness. This oversight could affect the overall efficiency and balance of the network, as not all content is equally suitable for caching based on these unconsidered attributes.

Reference [146] introduces a genetic-algorithm-based collaborative caching scheme (CCS) for energy optimization in the Content-Centric Internet of Things (C2IoT). By integrating both on-path and off-path strategies, the CCS aims to enhance the efficiency of content requests across the network. Utilizing genetic algorithms, the scheme optimizes caching decisions with an emphasis on content popularity, historical user requests, and transmission delays, striving to reduce communication delays and overall energy consumption. This approach represents a significant step towards leveraging the potential of collaborative caching to improve network performance and sustainability. However, the methodology has its limitations. Despite facilitating collaboration among nodes for more strategic content placement, individual decision-making can lead to off-path caching that doesn't completely eliminate redundancy. The analysis overlooks the transient nature of IoT content, which can vary in relevance over short periods, and fails to account for other content characteristics that could influence caching efficiency. Furthermore, the assumption of a robust infrastructure, with routers strategically located near IoT sensors, might not hold in many real-world scenarios, where network topology and accessibility challenges can significantly impact the feasibility of the proposed caching strategy.

Reference [147] introduces Trusted Caching Nodes (TCNs), innovative entities designed to revolutionize content caching and security within the Industrial Internet of Things (IIoT). These TCNs are engineered to fulfill dual purposes: efficiently caching frequently requested content and enhancing the security framework of the IIoT. By formulating a joint optimization problem, the study seeks to strategically position TCNs and delineate secure content distribution pathways, leveraging metrics such as content popularity, energy implications of caching and data transmission, and the costs associated with security breaches. However, the methodology has its drawbacks. Primarily, the necessity to execute the optimization model for every content retrieval request poses potential challenges, including significant energy expenditure and delayed response times, issues that starkly contrast with the inherent objectives of caching. Additionally, while the concept of Security Damage Cost is pivotal, the paper lacks a clear exposition on how this metric is quantitatively assessed. Moreover, the transient nature of IoT content, which demands dynamic and flexible caching strategies, appears to be overlooked in their approach, raising questions about the model's adaptability to the rapidly changing landscape of IIoT data needs.

The work presented in [148] introduces the Lifetime-based Cooperative Caching (LCC) scheme, a novel strategy tailored for ICN within IoT environments. This scheme adeptly leverages IoT data's lifetime alongside user request rates to inform caching decisions across the network, optimizing the use of in-network caching capabilities to extend the sleep mode durations of IoT devices like sensors and actuators, thereby enhancing energy conservation. Through a cooperative model, nodes such as content routers and base stations collaborate, employing a dynamic caching threshold that adjusts in response to fluctuations in content request rates. The decision to cache content considers factors such as the data's lifetime, its

popularity, and the energy consumption implications of caching. However, the methodology does not clearly outline how energy consumption is quantified. Despite its innovative approach, the LCC scheme falls short by omission of cache-aware routing mechanisms which restricts the ability to effectively utilize cached content on off-path nodes, potentially limiting the efficiency of the network's content delivery system. Furthermore, this approach raises concerns about the redundancy of cached content across the network. Without explicit strategies to mitigate such redundancy, there's a risk of decreasing the diversity and effectiveness of cached content, which could inadvertently increase energy consumption and reduce the overall efficiency of caching in the IoT ecosystem.

Authors in [149] propose an innovative approach to augment caching efficiency within IoT networks by leveraging the popularity of IoT data, as delineated in "IoTCache: Toward Data-Driven Network Caching for Internet of Things." Employing a dual-faceted strategy, the method combines Deep Neural Networks (DNN) for forecasting data popularity with statistical techniques designed to address the cold start problem, which arises in scenarios lacking sufficient historical data. This predictive model aims to optimize network resources by intelligently caching content that is anticipated to be in high demand. However, the paper's primary reliance on content popularity as the sole metric for caching decisions may not fully capture the nuanced and dynamic nature of IoT environments. IoT content often exhibits a transient relevance, with its value and applicability diminishing over time. In addition, in IoT settings, the popularity of data often correlates not with individual content items but with categories defined by sensor types, specific locations, or content genres. These broader context factors drive demand patterns, suggesting that a more effective caching strategy would consider the collective popularity within these contextual groupings rather than isolating single content items. Investigating popularity within this broader, category-based context could yield a more accurate and efficient caching mechanism, better aligned with the unique dynamics of IoT data generation and consumption.

Building on the need for efficient caching mechanism for IoT, [150], propose a sophisticated strategy that employs AI-based techniques, including collaborative filtering and k-means clustering, to optimize content caching at edge nodes. This strategy is designed to manage traffic efficiently, minimizing the load on cloud databases. By analyzing user request patterns and leveraging the predicted popularity of content, the system aims to intelligently decide which content to cache. The goal is to enhance the user experience by improving the cache hit rate, reducing the average hop count, and minimizing the Content Retrieval Delay. Despite these advancements, the paper outlines a collaboration mechanism for sharing information among nodes to facilitate intelligent caching decisions, yet it stops short of addressing the optimization of content placement across the network. This means finding the most strategic node for caching specific content, based on various network parameters and content characteristics, isn't directly tackled. Additionally, the proposed model lacks a mechanism for redirecting requests to nodes where content is already cached, which could further reduce retrieval times and network traffic. The methodology also does not fully account for the unique characteristics of IoT data, such as variability in data generation rates, sizes, and the temporal relevance of content, which are critical for truly optimizing caching in IoT environments.

Research in [151] introduces an approach to enhancing content caching strategies within IoT networks, leveraging the combined strengths of LSTM models and ensemble learning tailored for mobile/multiaccess edge computing (MEC) environments. This innovative methodology predicts content popularity across varied user demographics using LSTM models, while ensemble learning amalgamates these diverse predictions to establish a unified,

network-wide caching strategy. This predictive framework is designed to proactively cache content at MEC servers, aiming to optimize network resources by reducing latency and adapting to user demand in real-time. However, the application of user grouping and content popularity predictions in IoT contexts, where the range and type of IoT sensors within a MEC server's coverage may be limited, presents potential limitations. Unlike multimedia content, which benefits from diverse categories and user interests, the IoT environment often features a narrower spectrum of content types at least for the types that may absorb the interests of many users, making the direct application of such a user-aware caching strategy less effective. Furthermore, this approach does not fully account for the unique characteristics of IoT data or the dynamic nature of user interactions within these networks, raising concerns about its practical efficacy in real-world IoT scenarios.

Research in [152] optimizes the caching landscape within IoT networks through a fog caching strategy, adeptly engineered to cater to the unique demands of IoT applications. This methodological innovation stands out by weaving together three critical elements: the agility of adjustable clusters that tailor caching resources to current demands, the strategic near-path content allocation to enhance cache hits across varied user paths, and a holistic "Full-Time Caching" approach. This latter approach combines reactive caching, triggered by immediate content requests, with proactive caching in off-peak hours, aiming to preload the cache with content anticipated to be in high demand. Central to this strategy is the reliance on content popularity, serving as the primary metric guiding which content merits caching, leveraging predictions to preemptively meet user needs. However, the schema's embrace of off-peak caching raises pragmatic concerns within the perpetually evolving IoT environment. The assumption that content selected during quieter periods will retain its relevance and demand until the subsequent off-peak cycle overlooks the IoT domain's intrinsic dynamism. Moreover, the strategy's focus sidesteps the nuanced characteristics specific to IoT content, such as data generation rates, size, and contextual relevance, which are pivotal in tailoring a caching strategy that mirrors the IoT ecosystem's complexity and variability.

The work presented in [153] introduces a novel edge caching strategy that employs a cluster-based efficient caching mechanism within an ICN framework, tailored for edge devices in data-intensive IoT applications. By utilizing a clustering approach, the system organizes network devices into clusters, each led by a designated cluster head responsible for caching decisions. This architecture allows the cluster head to have a comprehensive view of the cache status within its cluster, enabling decisions that enhance content availability and network efficiency based on content popularity. Furthermore, clusters are designed to collaborate, enhancing the network's ability to respond to requests efficiently. However, the proposed strategy has its shortcomings, mainly in its evaluation of potential caching nodes. The cluster head's decisions focus predominantly on available storage space, sidelining other crucial node-specific metrics such as location, processing capabilities, and more. Moreover, the study does not address the transient nature of data in IoT environments, which could significantly impact the relevance and efficiency of the cached content over time. These oversights suggest areas for future research and refinement to fully harness the potential of ICN frameworks in edge computing for IoT applications.

Research in [154] optimizes content caching strategies for IoT networks by introducing an approach that prioritizes data freshness and popularity. The authors propose a method that formulates the caching problem as an integer linear programming (ILP) problem, subsequently modeled as a Markov decision process (MDP). To navigate the complexities of this optimization problem, a deep reinforcement learning algorithm (DRLA) is employed, utilizing DNNs to approximate state-action values and learn effective caching policies. The main criteria

for caching decisions are centered around Data Freshness and Request Probability, with IoT gateways serving as the pivotal nodes for caching and executing the DNN and reinforcement learning mechanisms. However, the paper presents notable shortcomings, particularly in its exploration of gateway capabilities. It does not specify the supportable number of nodes per gateway, raising concerns about the sufficiency of gateway storage for effective content caching, especially in densely networked environments. Moreover, the absence of a mechanism for node cooperation might lead to inefficiencies, with some nodes being overwhelmed and others underutilized. Incorporating cooperative strategies could not only alleviate load imbalances but also enhance overall network performance by leveraging caching for optimal load distribution.

Table 2.1. provides a comprehensive comparison between various related works in the context of caching strategies employed in IoT networks. This comparison is structured around six pivotal metrics, each reflecting a distinct aspect of caching dynamics and decision-making criteria.

References	Transient Content Awareness	Content Metrics Consideration	Node Capability Assessment	Topology-Based Caching	Hierarchical Caching Strategy	Smart Routing for Cache Access
141	Yes	No	No	No	No	No
142, 143, 148, 154	Yes	Yes	No	No	No	No
144	Yes	Yes	No	No	Yes	No
140	Yes	Yes	Yes	No	No	No
145	Yes	No	No	No	Yes	No
146, 147, 152	No	Yes	No	Yes	No	Yes
149, 150, 151	No	Yes	No	No	No	No
153	No	Yes	No	Yes	Yes	Yes

Table 2.1. Comparative Analysis of Caching Strategies in IoT Networks Across Key Metrics

- **Transient Content Awareness** evaluates whether the approach accounts for the ephemeral nature of IoT data, emphasizing strategies that adapt to changing content relevance over time.
- **Content Metrics Consideration** assesses if caching decisions factor in specific attributes of content, such as popularity and size, to optimize their caching strategy.
- **Node Capability Assessment** explores whether the strategies consider the varying capacities of individual nodes, including storage and computational power, which can significantly influence caching efficacy.
- **Topology-Based Caching** investigates the incorporation of network topology in caching decisions, potentially leveraging spatial relationships for more efficient

content distribution.

- **Hierarchical Caching Strategy** examines the presence of a coordinated mechanism overseeing caching across the network, aimed at reducing redundancy and enhancing overall cache utilization.
- **Smart Routing for Cache Access** scrutinizes the integration of intelligent routing mechanisms to direct requests towards cached content efficiently, thereby improving cache hit rates.

Research Gap

The investigation, as delineated so far through the comparative analysis, uncovers significant research gaps in existing caching policies and algorithms within IoT networks. Predominantly, the focus has been on caching strategies centered around individual nodes, largely overlooking the collective network environment and the potential for inter-node cooperation. This singular approach can precipitate redundancy, where the same content is cached multiple times across the network, and lead to suboptimal cache placements that do not fully leverage the network's topology or the nodes' heterogeneous capabilities. Furthermore, the absence of smart routing mechanisms becomes increasingly detrimental in larger scenarios with numerous potential paths for a request to traverse, resulting in a diminished likelihood of cache hits. Moreover, there is a clear necessity for a holistic approach that harmonizes content characteristics with node capabilities, ensuring that caching decisions are not only informed by what is being cached but also by where and by whom within the network. Addressing these gaps promises significant advancements in the efficiency and effectiveness of caching strategies in the complex and dynamic landscapes of IoT networks.

2.5.2. Advancements in Caching for IoV Systems

In this subsection, we focus on the specific advancements in caching strategies tailored for IoV systems. The purpose here is to examine how caching has adapted to meet the unique requirements of IoV, including high mobility environments and the need for real-time data processing. Through an overview of related work, we will highlight innovative caching approaches that have been developed to improve IoV performance, such as reducing latency, enhancing location-aware content delivery, and supporting efficient data dissemination. This exploration underscores the critical role of caching in advancing IoV technologies and facilitating smarter, safer vehicular networks.

The work in [155] introduces a novel edge caching strategy, the Cooperative Caching Strategy with Content Request Prediction (CCCRP), which utilizes vehicles and RSUs for caching content anticipated to be requested by vehicles. A distinctive method is proposed, leveraging vehicle clustering via K-means to streamline content request and transmission processes. It harnesses LSTM networks for predicting vehicle content requests based on historical data, followed by employing reinforcement learning to optimize the objective function for content placement, thus aiming to enhance QoS for vehicle requests. However, the study has its limitations, notably the omission of privacy considerations in the popularity prediction process. Furthermore, it does not differentiate between content types such as safety and infotainment in the IoV environment which each possess unique characteristics and may influence caching strategies and priorities differently.

In [156], an edge caching strategy employs a methodology to enhance content delivery within vehicular networks. The authors introduce a scheme focusing on the dynamic and cooperative caching of content in both RSUs and vehicles, based on an analysis of vehicular content requests. This analysis considers variables such as content access patterns, vehicle

velocity, and road traffic density to establish a model for fetching content replicas efficiently when requested by moving vehicles. The proposed method optimizes the use of RSUs and vehicular storage to facilitate low-latency downloads directly from the nearest RSU or, if necessary, through content sharing among vehicles, aiming to reduce content retrieval delays. However, the study's approach to caching decisions primarily centers on content popularity, overlooking the potential impact of other content characteristics. Furthermore, it does not differentiate between various content types prevalent in the IoV environment, such as safety-related and infotainment content, each bearing unique significance and requirements that could influence caching strategies and network performance.

Authors in [157] introduce a method that optimizes the content caching strategy for Vehicular Named Data Networks (VNDN) by methodically categorizing data into three distinct types to address the dynamic nature of IoV environments efficiently. The authors propose a novel approach that intelligently divides information into emergency safety messages, traffic efficiency data, and infotainment content. Safety messages, which may be periodic or triggered by emergencies, are prioritized differently from traffic data that aids in route planning, potentially reducing travel times and fuel consumption. Infotainment content, encompassing multimedia data, is cached based on its popularity. They suggest the use of integer programming to make informed caching decisions that could enhance network performance by tailoring the caching strategy to the specific characteristics of each data type. However, the paper stops short of detailing a mechanism for predicting the popularity of content, a crucial aspect for proactive caching in the highly dynamic IoV landscape. Additionally, there seems to be a lack of emphasis on the potential cooperation between RSUs, which could further leverage the distributed nature of caches. This omission suggests that the benefits of their caching strategy might be constrained to the local areas where the cached content resides, potentially limiting the overall efficacy of the proposed approach in a broader IoV context.

In [158], an edge caching strategy employs a hierarchical network architecture optimized for the dynamic and resource-constrained IoV environment. The authors introduce a cooperative caching scheme that divides vehicles into clusters, selecting one vehicle per cluster as the cluster head for cache management. Communication within a cluster is facilitated exclusively via the cluster head. When a cluster member requests content not available in its local cache, the request is directed to the cluster head, which then queries the entire cluster. Should the content be cached by any cluster member, it is promptly sent to the requester. Otherwise, the cluster head might extend the request to neighboring clusters. Upon receiving new content, a vehicle decides whether to cache it based on available space and a comparison of caching probabilities between new and existing content, utilizing metrics such as content popularity and future ratings predicted through Non-negative Matrix Factorization (NMF). Despite its innovative approach, the study's reliance on request counts to calculate content popularity might not effectively indicate future content relevance in the highly dynamic IoV context. Additionally, the methodology lacks a privacy-preserving mechanism for predicting future ratings, raising concerns about user privacy in the proposed caching strategy.

The work presented in [159] introduces a novel edge caching strategy employing digital twin technology and deep learning for optimizing content delivery in the IoV. The method suggests a data caching mechanism at the edge, using social data where content is categorized into types, each defined by unique average size and maximum acceptable delay. Besides leveraging RSUs for storage, contents are pre-stored in smart vehicles, considering the social relationship among them, based on content compatibility, information exchange rate, traffic

situation, and vehicle density. This dynamic social interaction is modeled using a digital twin, constructed with vehicle data collected by RSUs, enabling an adaptive caching resource allocation that enhances the content delivery process. However, the study's approach to quantifying content popularity and the intricacies of user preference similarities raise questions. It notably omits a detailed method for quantifying content popularity—a critical factor in cache management. Moreover, the reliance on user preference similarities and social interaction data raises potential privacy concerns, suggesting a need for comprehensive privacy-preserving measures within the system's design.

Research in [160] optimizes the utilization of in-network caching within 5G-enabled vehicular networks, laying out a detailed framework where both vehicles and small cell base stations (SBSs) partake in content caching. It dissects service content into manageable chunks, allowing vehicles to process one chunk while simultaneously requesting subsequent ones. This method strategically mitigates user-perceived delays by enabling an immediate access point to cached content within a vehicle's proximity, or through SBSs, before resorting to a macro base station (MBS) in the absence of locally cached content. The entire operational blueprint is formulated as an optimization problem, aiming at elevating the user experience by curtailing delays through proactive content caching and distribution strategies. However, the scheme's efficacy could potentially be undermined by the inherent complexities in content segmentation and caching distribution. If the segmentation isn't meticulously planned, it might result in significant inefficiencies, particularly if various segments are scattered across different locations, thus impeding swift content assembly. Moreover, the approach presupposes an advanced, albeit intricate, mechanism for predicting content popularity down to individual segments. The current methodology does not thoroughly address the need for such a refined popularity prediction mechanism, which is crucial for optimizing the caching process and ensuring the model's overall success.

Reference [161] proposes a user-experience-based caching method tailored for the IoV, with an emphasis on short videos due to their popularity and high sensitivity to download delays. The central hypothesis posits that a user's interest in a specific file likely indicates an interest in similar content, leading to the utilization of file class popularity distribution for caching decisions. An QoE model underpins this approach, predicated on the ratio of files matching user interest to the total files within an RSU. To navigate the complexities of optimizing this user experience model, the strategy leverages reinforcement learning for making informed caching decisions in RSUs to enhance QoE. However, the research's focus on small files may limit its applicability across varied content types, constraining its utility in broader IoV scenarios. Additionally, the absence of detailed criteria for content classification into classes raises concerns regarding the consistency of interest and request patterns within each class, potentially affecting the accuracy and efficiency of the caching mechanism.

In [162], the authors outline a task-based architecture for enhancing content caching within the IoV, featuring a process divided into content popularity prediction using regression and deep learning techniques, strategic content placement within cache leveraging artificial intelligence for optimization, and streamlined content retrieval that involves intelligent allocation of communication resources for efficient delivery. This approach focuses on utilizing RSUs as pivotal nodes for caching, effectively reducing latency and improving data access speeds for vehicular networks. However, the study presents limitations in its scope, notably overlooking the diverse nature of content types within the IoV ecosystem and their specific caching needs. It lacks a detailed examination of cooperative caching strategies that could exploit the interconnected nature of RSUs for enhanced content availability. Additionally, the research primarily emphasizes content popularity while giving minimal consideration to other

crucial content attributes, such as size, which could significantly influence caching decisions and overall network performance.

The work in [163] proposes a proactive caching scheme for the IoV networks using an Information-Centric Network (ICN) approach. Given the limited computational and storage capabilities of edge devices, a scheme is introduced that employs a Markov process to model real-time vehicle status and interactions with other vehicles and RSUs. This model informs the creation of an optimal caching decision, aimed at maximizing network performance and reducing transmission delay. It takes into account the dynamism of topology changes due to vehicle mobility and potential link failures. However, this approach does not consider the nature of the content within the IoV ecosystem, which could affect the relevance and utility of cached data. Moreover, the efficiency of the proposed method hinges on access to extensive private information, such as vehicle trajectories, raising concerns about data privacy and the feasibility of obtaining such detailed information in practice. These shortcomings suggest that while the proposed model offers a structured approach to edge caching in IoV, there is room for improvement, particularly in addressing privacy concerns and content relevance.

Research in [164] optimizes edge caching in IoV networks through two distinct strategies: Ensemble Learning-Driven Edge Caching (ELDEC) and Meta-based Ensemble Learning-Driven Edge Caching (MELDEC). Employing meta and ensemble learning, these strategies aim to refine content popularity prediction and enhance cache content placement, directly impacting the network's wireless performance by deriving closed-form expressions for outage probability and finite SNR diversity gain. Despite the strides in managing network efficiency and predictive accuracy, the study overlooks critical aspects such as the integration of safety content considerations and the establishment of privacy mechanisms within the popularity prediction process. These omissions highlight potential areas for further refinement, particularly in addressing the nuanced demands of IoV environments where safety and privacy are paramount.

In [165], the challenge of efficient computation offloading in 6G-enabled networks is addressed. The authors lay out a framework, dubbed Computation Offloading with Demand prediction and Reinforcement learning (CODR), which leverages a Spatial-Temporal Graph Neural Network (STGNN) for anticipating demand, subsequently informing a caching decision process via the simplex algorithm. To finalize their strategy, a computation offloading approach using the twin delayed deterministic policy gradient (TD3) method is deployed to ascertain the most effective offloading pathway. Despite the thoroughness of the methodology, the work is not without its limitations, notably the omission of privacy considerations in the data-sharing process essential for demand prediction and caching decisions. Furthermore, the adaptability of CODR to the IoV environment's rapid changes, such as varying network conditions and vehicular mobility, is not sufficiently explored.

Reference [166] delves into the escalating data traffic challenges within IoV networks, putting forward a dual-strategy aimed at bolstering QoS. The approach kicks off with a spatial-temporal correlation method to sharpen the prediction of content popularity, utilizing historical data from RSUs. Building on this foundation, the study pioneers a Multi-Agent Reinforcement Learning (MARL) caching strategy. In this setup, each RSU functions as an independent agent, navigating through decisions and updates in action to heighten the caching efficiency amidst the IoV's dynamic landscape. Despite its strengths, the methodology exhibits notable limitations. Chief among these is the autonomy in decision-making by individual RSUs, which does not fully exploit the potential for collective intelligence and resource sharing among RSUs. This isolated approach may not be entirely conducive in the

high-mobility IoV context, where vehicles frequently transition across RSU coverage zones, potentially benefitting from a more integrated caching strategy that leverages cached content across neighboring RSUs.

The study in [167] addresses the challenge of content caching in vehicular networks (VNs) by introducing a Mobility-aware Proactive edge Caching scheme based on Federated learning (MPCF). This approach integrates a Context-aware Adversarial AutoEncoder (C-AAE) and a mobility-aware cache replacement policy to predict and adapt to the changing popularity of content due to vehicle mobility, emphasizing privacy preservation and communication efficiency. The MPCF scheme operationalizes this through four pivotal components: content popularity prediction, vehicle selection for federated learning, model aggregation, and cache replacement, enabling vehicles to collaboratively refine a global model for proactive content caching. However, the focus predominantly on content popularity might overlook other critical content attributes, such as size or the specific relevance of safety-related information, which can be vital for an exhaustive and nuanced caching strategy.

In [168], researchers present an edge caching approach for the IoV utilizing Multi-Agent Deep Reinforcement Learning (ECMADRL), wherein each moving vehicle operates as an independent agent. These agents are tasked with making dynamic decisions concerning the caching and accessing of content, guided by real-time environmental variables such as content demand, size, and the vehicle's caching capacity. This model's core lies in enabling vehicles to adaptively respond to changing conditions, leveraging deep reinforcement learning for decision-making processes. While the approach facilitates adaptive and decentralized caching decisions, potential challenges include the risk of redundant caching, as individual vehicles may independently decide to cache the same content without full visibility of the network's overall caching status. Additionally, the system's reliance on vehicles' local observations to calculate content popularity could lead to inaccuracies, given the limited scope of data available to each vehicle. These aspects may impact the overall efficiency and scalability of the proposed caching strategy within the dynamic IoV environment.

The study in [169] explores a caching policy aimed at enhancing content retrieval efficiency in autonomous connected vehicles through C-V2X communications. It introduces utility functions for both vehicles and RSUs, calculating the former based on the distance to RSUs to minimize retrieval delay and the latter on content popularity and the anticipated duration vehicles remain within coverage, employing the Gale-Shapley stable matching algorithm for strategic cache allocation. Despite its detailed approach to utility-based caching, the methodology may not preemptively account for the immediate content requests of vehicles upon entering RSU coverage areas, suggesting a reactive rather than a fully proactive caching strategy. Additionally, the potential for coverage area overlap between RSUs raises concerns about redundant content caching, indicating areas where further refinement could enhance network-wide caching efficiency and reduce resource wastage.

Table 2.2. presents a detailed comparison of the current research landscape concerning edge caching strategies within the IoV. This comparative analysis is structured around six meticulously chosen metrics, each shedding light on the diverse approaches and considerations in the development of caching mechanisms.

- **Safety and Traffic Content Tuning:** Investigates whether caching mechanisms are specifically optimized for critical safety and traffic information, recognizing its paramount importance in vehicular networks.
- **Infotainment Content Adaptation:** Assesses the tuning of caching strategies for infotainment content, which has distinct requirements compared to safety-related data.

- **Content Characteristics Consideration:** Evaluates the extent to which caching decisions take into account a broader spectrum of content characteristics beyond mere popularity, including spatiotemporal attributes.
- **Popularity Prediction Mechanism:** Distinguishes between methodologies that rely on static request counting and those employing dynamic approaches, such as machine learning, to forecast content popularity.
- **Privacy Preservation Mechanism:** Identifies efforts to incorporate privacy-preserving mechanisms within the caching strategy, essential for protecting user data during collection and sharing.
- **Collaborative Caching Strategy:** Explores the potential for and implementation of collaborative mechanisms among nodes for optimized cache retrieval and decision-making, aiming for a more globally optimal caching strategy.

References	Safety and Traffic Content Tuning	Infotainment Content Adaptation	Content Characteristics Consideration	Popularity Prediction Mechanism	Privacy Preservation Mechanism	Collaborative Caching Strategy
155, 156	No	No	No	Yes	No	Yes
157	Yes	Yes	Yes	No	No	No
158, 159	No	No	Yes	No	No	Yes
160	No	Yes	No	No	No	Yes
161	No	Yes	Yes	Yes	No	No
162, 164, 166	No	No	No	Yes	No	No
167	No	No	No	Yes	Yes	No
163, 168	No	No	Yes	No	No	No
165, 169	No	Yes	No	Yes	No	No
168	No	No	Yes	No	Yes	Yes

Table 2.2 Comparative Evaluation of Edge Caching Strategies in IoV: Focusing on Content Dynamics and Collaborative Mechanisms

Research Gap Explanation

The investigation into edge caching strategies in IoV reveals a predominant focus on popularity as the principal criterion for caching decisions, often neglecting the rich spatiotemporal characteristics and the varied nature of content across different vehicular applications. While advancements have been made in predicting content popularity, including

efforts towards privacy preservation, there remains a substantial gap in the strategic selection of nodes for participation in the learning process and the formulation of policies to achieve high-quality models. Furthermore, the prevailing approach to making caching decisions is characteristically local, overlooking the benefits of a globally optimized strategy. This localized decision-making process fails to fully capitalize on the potential of collaborative efforts among nodes, which could significantly enhance the efficiency and effectiveness of caching mechanisms. Moreover, the distinct demands of various content types, from safety messages to infotainment, necessitate unique caching strategies tailored to their specific requirements. Addressing these gaps by considering the inherent nature and spatiotemporal characteristics of IoV content, and fostering cooperation among nodes in caching decisions, is crucial for developing a comprehensive and effective edge caching model. This model would not only improve cache hit rates but also ensure that the diverse needs of vehicular network users are met efficiently and securely.

2.5.3. Integrating UAVs with Caching Strategies: A New Horizon

This subsection explores the cutting-edge integration of UAVs with caching strategies, marking a new horizon in network optimization and content delivery. The aim is to shed light on how UAVs, when combined with caching mechanisms, offer novel solutions to extend network coverage, improve content delivery in remote or underserved areas, and enhance the resilience of communication systems. By reviewing related works, we will identify the challenges and opportunities presented by this integration, illustrating its potential to redefine traditional network paradigms and introduce more flexible, efficient, and scalable solutions. This analysis highlights the innovative intersection of UAV technology and caching strategies, paving the way for future advancements in wireless communication networks.

The work presented in [170] introduces an approach to enhance MEC through ultrareliable caching for edge-enabled UAV networks, employing drones as on-demand nodes for efficient caching. This study leverages a neural-blockchain model aiming to improve reliability in communications within UAV-supported edge networks. The methodology emphasizes the use of blockchain for decentralized, secure data transactions and neural networks for optimizing caching decisions based on network dynamics. However, the approach appears to overlook potential benefits of direct cooperation among UAVs for sharing and retrieving cached content, which could further reduce latency and network load. Additionally, the strategy primarily focuses on network parameters to determine the caching of content, without explicitly incorporating content-specific metrics such as popularity or user preferences, which could enhance the efficiency and relevance of cached data. The absence of a detailed mechanism for content selection based on such metrics may limit the adaptability and effectiveness of the caching strategy in dynamic network environments with varying content demand patterns.

The work in [171] explores the integration of UAVs, VANETs, and caching mechanisms to address the challenges of content delivery in intelligent transportation systems. This study centers around the development of a system where a UAV autonomously makes caching decisions based on the popularity and size of content, along with its current cache status, to optimally serve vehicles on the road. The approach involves formulating the problem as a Mixed Integer Non-Linear Problem (MINLP) and subsequently as an MDP, which is solved using Proximal Policy Optimization (PPO), highlighting the UAV's trajectory planning, cache management, and radio resource allocation strategies. Despite the outlined system's potential

to improve content delivery efficiency, several shortcomings are evident. The absence of a dynamic popularity prediction mechanism limits the system's ability to proactively cache content based on real-time demand changes. Moreover, the study primarily examines the operation of a single UAV, which may not reflect the complexities and potential benefits of multiple UAVs working in coordination. Additionally, the practicality of implementing computationally intensive algorithms like PPO on UAVs with limited computational resources is not thoroughly assessed, which raises questions about the scalability and real-world applicability of the proposed approach in diverse operational environments.

The work in [172] introduces FLchain, integrating MEC with Federated Learning (FL) and blockchain technologies to address issues of data privacy and centralization, aiming to bolster security and privacy within intelligent MEC networks. This framework is particularly emphasized for its role in facilitating decentralized, secure, and privacy-preserving communications across edge computing applications, including UAV networks, edge data sharing, and content caching. Despite the breadth of its applications, the paper stops short of detailing specific caching mechanisms, proposing efficient UAV placement strategies, or considering the energy constraints critical for UAV operations. Furthermore, while it outlines the conceptual underpinnings and potential benefits of FLchain, it does not delve into the operational metrics or decision-making criteria for content caching, leaving a gap in the practical implementation of its proposed architecture.

The work presented in [173] focuses on improving UAV-assisted wireless caching networks (WCN) through the implementation of Non-Orthogonal Multiple Access (NOMA). It leverages UAVs for mobile caching, employs NOMA for serving multiple users over the same spectrum, and introduces algorithms for efficient power allocation and user clustering, notably the p -K-means algorithm for UAV deployment optimization based on user clustering. This work is centered around a cross-layer resource allocation strategy to enhance system performance, particularly by maximizing the system hit probability through strategic file caching and power distribution decisions. However, the approach exhibits several limitations. Primarily, it does not account for user interests or behaviors when clustering, relying solely on proximity and channel gain, which might not accurately reflect the dynamic and localized content demand within specific UAV coverage areas. Furthermore, the centralized content popularity prediction by the base station, without considering individual UAV areas and potential user mobility, may not fully align with the immediate needs and interests of users under specific UAVs. Notably, the study overlooks the energy constraints of UAVs, a critical aspect in designing UAV-assisted networks, as these limitations directly impact UAVs' operational capacity and, consequently, the efficiency of content delivery and network performance. This omission underscores a gap in ensuring the sustainability and long-term viability of the proposed network configuration under real-world conditions where energy resources and user demands are variable and unpredictable.

The study in [174] introduces an approach that blends cache-enabling UAV cellular networks with NOMA technology to offload wireless backhaul link traffic and enhance network performance by caching popular content on a mobile UAV base station. Central to this exploration is the development of an MDP framework for the long-term optimization of caching placement and resource allocation, tackled through Q-learning algorithms with the UAV acting as a decision-making agent. This strategy is poised to meet the surging demand for high data rates and low access delay in multimedia content delivery. Despite its thorough exploration of UAV-based caching and resource allocation, the research does not delve into the specifics of UAV placement strategies, the collaborative potential between multiple UAVs, or a dynamic method for determining the popularity of content based on real-time user

request patterns. The focus is predominantly on the operational optimization of a single UAV without considering the variable characteristics of the content itself or the potential for enhancing network coverage and efficiency through strategic UAV positioning and multi-UAV collaboration.

Research in [175] optimizes the deployment and operation of UAVs within mobile edge networks (MEN) by focusing on content caching strategies designed to improve the network's ability to manage data traffic and enhance the QoE for users. The QoE is calculated based on transmission delays and their impact on user satisfaction, modeled through a DRL framework. This approach incorporates human-centric features, such as user mobility and request patterns, into the DRL algorithm to predict the most effective locations for UAVs and the optimal content to cache, addressing the issue by formulating it as an MDP. Despite its contributions to improving MEN efficiency, the study has certain areas that could benefit from further exploration. It does not detail mechanisms for preserving user privacy when integrating human-centric data, nor does it address the potential energy limitations associated with UAVs, which could significantly impact their operational time and effectiveness. Furthermore, the possibility of UAV collaboration for content sharing, which could potentially enhance network performance, is not explored, missing an opportunity to further optimize content delivery and caching in MENs.

Research in [176] optimizes the use of UAVs for mobile edge caching within wireless communication systems, leveraging Latent Dirichlet Allocation (LDA) for enhancing caching strategies through an analysis of user request preferences. This approach integrates k-means clustering to effectively classify users, aiding in the strategic deployment of UAVs according to user distribution and preferences. Centralized analysis on a cloud server underpins the caching decisions, focusing on content popularity and anticipated demand among users. However, the study does not delineate the division of user preferences across distinct UAV coverage areas, leading to a potentially generalized caching strategy that may not align with localized user demands. Moreover, the reliance on historical browsing data for preference analysis without addressing privacy concerns or implementing any privacy-preserving mechanisms could pose significant privacy risks. This absence of specificity in user preference localization and the lack of privacy safeguards mark critical areas for improvement in the proposed system.

The study in [177] optimizes the deployment of UAVs in device-to-device (D2D) communication networks through a framework that strategically manages both UAV flight trajectory and caching placement. It tackles the challenge of maximizing cache utility within these networks by formulating and decomposing a mixed integer nonlinear programming problem into three solvable sub-problems: user terminal (UT) caching placement, UAV trajectory optimization, and UAV caching strategy. This decomposition approach utilizes a many-to-many swap matching algorithm, approximate convex optimization, and dynamic programming to address each sub-problem respectively. The role of Scalable Video Coding (SVC) is pivotal in the proposed strategy, enabling variable content quality levels to be efficiently cached and delivered based on network demand. However, the research does not adequately address several practical considerations. Fluctuations in content popularity over time are overlooked, potentially reducing the efficacy of the caching strategy in dynamic network environments. The absence of a privacy-preserving mechanism for data sharing between UAVs and UTs raises concerns regarding user data confidentiality. Additionally, the independent solution of decomposed sub-problems may lead to potential misalignments, such as sub-optimal UAV positioning relative to its cache content, impacting overall network performance. While the study acknowledges the importance of content popularity in caching

decisions, it fails to differentiate between the demand for various quality levels of the same content, which could lead to inefficient storage use. Moreover, by focusing primarily on the caching capabilities of UAVs without fully integrating their communication facilitation roles, the proposed framework might not be practically justifiable considering the operational costs associated with UAV deployment.

Research in [178] optimizes UAV deployment in IoV networks, focusing on caching mechanisms to improve high-bandwidth content delivery. The study formulates a Joint Caching and Trajectory Optimization (JCTO) problem, approached through a deep supervised learning method. This method encompasses two strategies: Offline Optimization, which clusters vehicles to manage computational complexity and optimizes content placement and UAV trajectory, and Online Decision-Making with Deep Learning, allowing for adaptive real-time adjustments based on dynamic network conditions. While this approach contributes to the efficient allocation of UAV resources in vehicular networks, it does not fully address several potential challenges. Specifically, the study does not differentiate between content types, such as safety and infotainment, each with distinct requirements. Moreover, the reliance on offline optimization to train the online model may not adequately capture the IoV's dynamics, as the real-time model's effectiveness is inherently limited by the scope of the offline data it learns from. Additionally, the method's focus on proximal vehicle locations without considering their trajectories or future movements may reduce the effectiveness of UAV deployment in the highly dynamic IoV environment, potentially overlooking more efficient content delivery paths that account for vehicular mobility patterns.

Research in [179] optimizes cache placement strategies within UAV-assisted networks incorporating D2D communications, focusing on enhancing system performance through improved cache hit probability and trajectory optimization in both static and dynamic UAV deployments. This optimization aims at intelligently distributing content across the network, leveraging UAVs for content delivery and D2D for localized sharing, thereby alleviating the load on the backhaul network and enhancing user access to popular content. Despite its contributions, the study presents limitations such as the absence of a detailed methodology for predicting content popularity, which is crucial for implementing the probabilistic caching effectively. This gap raises questions about the practical execution of cache placement strategies, particularly in dynamic network environments where content popularity can fluctuate rapidly. Additionally, the analysis does not account for the redundancy in caching, particularly in scenarios where nearby devices cache identical content, potentially leading to inefficient use of limited storage resources. Another significant oversight is the neglect of UAV power constraints and battery dependency, which are critical factors in the practical deployment of UAV-assisted networks. Addressing these shortcomings could significantly enhance the feasibility and efficiency of the proposed caching strategies in real-world applications.

The work presented in [180] optimizes content distribution in 5G Heterogeneous Networks (HetNets) through a mechanism that strategically employs UAVs and the Internet of Connected Vehicles (IoCVs). The focus is on enhancing the QoE for vehicle users by formulating an optimization problem aimed at reducing transmission delays and applying a coalition game between UAVs and IoCVs to identify the most effective content distribution strategy. This approach intricately considers the utility functions of both UAVs and IoCVs, aiming to stimulate their participation in the content distribution process. However, the study does not delineate specific strategies for optimizing UAV battery usage, which is critical given their energy constraints. Moreover, it overlooks the elaboration of a robust UAV deployment strategy, crucial for the operational efficacy of such networks, and does not address

mechanisms to offload the computationally intensive tasks associated with the optimization problem from the UAVs, ensuring the sustainability of the optimization model's goals and quality. This leaves room for further exploration on improving UAV energy management, computational offloading for optimization tasks, and deployment strategies within HetNets environments.

In [181] authors introduce an optimization for content caching and UAV trajectory planning within the IoV ecosystem by introducing a framework that utilizes UAVs to enhance the delivery of delay-sensitive content amidst challenges of limited resources and high vehicular mobility. The approach, anchored on a content Caching scheme with Trajectory design through Differential Evolution and Deep Reinforcement Learning (CTDR), addresses energy-efficient content delivery by proposing a differential evolution-based content caching scheme and a trajectory optimization via Multi-Agent Proximal Policy Optimization (MAPPO). This dual-focus strategy aims at dynamically updating UAV caches and optimizing flight paths to reduce system energy consumption while ensuring high service quality. Despite the framework's contributions, the autonomy in UAV decision-making raises questions about the global network optimality, particularly in environments with highly mobile and transient vehicle requests. This model might face hurdles in achieving proactive caching that accurately predicts future demands, as it relies primarily on recent historical request data received by the UAV. The dynamic update mechanism, while responsive to recent requests, may not fully anticipate the rapid changes in vehicular networks, potentially affecting the system's overall effectiveness in real-time content delivery.

The work presented in [182] introduces a method for deploying in cache-enabled multi-UAV networks, aiming to enhance the QoE for ground users by reducing the average request delay. This is approached as a joint optimization problem, categorized as NP-hard, and addressed through the decomposition into two sub-problems: UAV deployment, leveraging a weighted K-means algorithm for user clustering based on geographical location and activity levels, and content placement, utilizing a Q-learning algorithm to dynamically adapt to changing content popularity and user demands. The study ambitiously attempts to navigate the complexities of dynamic content libraries and heterogeneous user activity without relying on static predictions of content popularity. However, it faces limitations, notably in its reliance on the user activity level, a metric determined by a probability assigned to each user, where the methodology for calculating these probabilities remains undefined. Moreover, the independence of UAVs in observing user requests could lead to suboptimal cache placements in scenarios where user mobility is significant, potentially resulting in users moving out of the UAV's coverage area before the system can effectively learn and adapt to their content preferences.

Research in [183] optimizes UAV-assisted cellular networks by focusing on the dual challenges of UAV deployment and content caching to enhance user QoE, evaluated via the Mean Opinion Score (MOS). This study proposes a joint optimization problem, tackling it through the decomposition into UAV deployment and caching placement. A swap matching-based algorithm is utilized for the strategic placement of UAVs, while a greedy algorithm is responsible for making caching decisions, considering factors like content popularity and size, albeit without a real-time prediction mechanism for the latter. Despite its comprehensive treatment of deployment and caching strategies aimed at minimizing transmission delays, the study overlooks several practical concerns. Notably, it does not address the power constraints of UAVs, which could significantly impact their operational longevity and service availability. Additionally, the paper lacks a detailed exploration of UAV collaboration possibilities and the integration of UAVs with existing cellular infrastructure beyond basic backhaul connectivity,

missing out on potential performance enhancements through cooperative dynamics and infrastructure synergy.

The study in [184] optimizes the caching strategy for UAVs in cellular networks, aiming to improve service timeliness and address the limitations of UAVs' finite storage capacities. It employs PPO, a reinforcement learning algorithm, to adaptively identify the best caching approach. A key contribution is the introduction of the Effective Age of Information (EAoI) metric, which assesses the freshness of information while accounting for packet errors during transmission. This approach enables dynamic adaptation to user demands and network conditions, enhancing the relevance and timeliness of the content delivered to users. However, the reliance on a reinforcement learning algorithm for real-time decision-making raises concerns about the computational and power demands on UAVs, which are typically resource-constrained. Additionally, the focus on content freshness may not be universally applicable across all types of content; for instance, many forms of multimedia content, like movies, do not require frequent updates to remain relevant to users, suggesting the proposed framework might have limited applicability in scenarios not driven by the need for up-to-the-minute information. The absence of a detailed strategy for dynamic UAV placement or considerations for optimizing communication quality, such as signal-to-interference-plus-noise ratio (SINR) and data rate adjustments, signifies missed opportunities to further enhance network performance and service quality.

Table 2.3. offers a systematic comparison across the spectrum of research focusing on UAV deployment integrated with caching strategies in next-generation networks. This analysis employs six critical metrics to distill the essence of each study's approach, highlighting the diversity and focus areas in current research.

- **Power Constraint Consideration:** Evaluates if studies account for the energy limitations of UAVs, which are crucial for prolonging operational durations and optimizing flight paths.
- **Caching Mechanism Detailing:** Assesses if the related work not only mentions caching but also elaborates on the specific mechanisms and procedures for implementing caching on UAVs.
- **Communication Metrics Awareness:** Investigates the consideration of communication metrics like path loss, data rate, and signal-to-noise ratios, which are fundamental for ensuring reliable UAV communication links.
- **UAV Collaborative Strategy:** Identifies the inclusion of strategies for UAV collaboration, essential for enhancing caching efficiency and decision-making processes through shared intelligence.
- **Integration with Terrestrial Infrastructure:** Explores whether UAV deployment strategies are conceptualized to function in synergy with ground-based infrastructures such as Base Stations (BS) and Road Side Units (RSU), acknowledging the value of a unified network ecosystem.
- **Decentralized Approach Adoption:** Examines the adoption of decentralized decision-making in UAV operations, crucial for scalability, robustness against failures, and minimizing cloud dependency.

Research Gap Explanation

The prevailing research on UAV deployment with caching often isolates UAV operational factors like energy efficiency and trajectory optimization, without considering them as integral parts of a holistic system encompassing terrestrial infrastructures, cloud computing, and sophisticated caching mechanisms. This siloed approach falls short of achieving enhanced QoS

across the entire network, underscoring the need for a more integrated methodology. A significant portion of the literature does not fully explore UAV collaboration nor its potential synergy with terrestrial networks; and where the integration with terrestrial infrastructure is acknowledged, it seldom progresses beyond conceptual proposals to detailed implementation strategies. Furthermore, the potential of decentralized network decision-making, which can avert the pitfalls of cloud reliance such as single points of failure and latency, is frequently overlooked. Moreover, there is a notable gap in addressing practical constraints such as UAV power limitations and the intricacies of content selection mechanisms, elements critical for the operational viability of UAV networks. Addressing these gaps necessitates a shift towards comprehensive strategies that not only improve network-wide QoS but also ensure the sustainable and efficient deployment of UAVs within the broader network infrastructure.

References	Power Constraint Consideration	Caching Mechanism Detailing	Communication Metrics Awareness	UAV Collaborative Strategy	Integration with Terrestrial Infrastructure	Decentralized Approach Adoption
170	No	No	Yes	No	Yes	Yes
171	Yes	Yes	Yes	No	No	Yes
172	No	No	No	Yes	Yes	Yes
173, 175	No	No	Yes	No	Yes	No
174	No	Yes	Yes	No	No	Yes
176, 180, 183	No	Yes	Yes	No	Yes	No
177	No	Yes	No	No	No	Yes
182	No	Yes	No	No	No	No
178	Yes	No	Yes	No	Yes	No
179	No	No	Yes	No	No	No
181	Yes	Yes	Yes	No	Yes	Yes
184	No	Yes	Yes	No	No	Yes

Table 2.3. Strategic Overview of UAV Deployment with Caching in Next-Generation Networks

3. A software-defined caching scheme for the Internet of Things

3.1 Introduction

The Internet of Things (IoT) is a paradigm where billions of heterogeneous devices connect and communicate. However, implementing new services using IoT is far from straightforward and faces many challenges, including, providing a connection between heterogeneous nodes and transient nature of data. An important aspect regarding IoT is the content-centric nature of IoT data, meaning that most applications designed for this ecosystem are generally interested in the content regardless of where it is located or by whom it is produced.

One of the approaches which can be used to mitigate the challenges related to the implementation of IoT is CCN [185], which is a new method that changes architecture and the concept of the current Internet. One of the main characteristics of CCN is naming the data, which makes each chunk a self-identifying unit. Retrieving data by their name instead of their location is another feature of CCN that can be used to support mobility, and finally, the most essential aspect of CCN is caching the data which can be exploited to solve many problems in the realm of IoT. Caching data can bring many benefits, for instance, similar to anycast in IPv6 [186], respond to requests from the closest node caching the data can bring many benefits such as reducing the immense access to constrained resource nodes, saving energy consumption and hence prolonging the life cycle of nodes. Caching can also improve availability by increasing the chance to access the data through the cached content, even if the original producer is not available. Additionally, in-network caching can be used to face the challenges that arise from big data and decrease the bandwidth usage by caching the proper contents on suitable devices.

In-network caching can be very useful in many use cases, for instance, consider the scenario of finding a parking space. The main characteristics of this scenario are the vastness of the environment, the dynamic nature of the environment, and a large number of users. Users in this scenario mainly need parking space information in their vicinity, so they can connect directly to the IoT network instead of the Internet and demanding the current parking space status in the area of their interest. The information gathered by IoT sensors will eventually be transmitted to the cloud for long-term storage and analysis. However, it is crucial that the data gathered by IoT devices regarding the location and status of parking spaces, be available at any time for users, even if because of intermittent connections, the cloud would not be available or sometimes because of QoS considerations the delay to retrieve data from the cloud will not be tolerable. Thus, caching data on IoT devices can be utilized to ensure data availability and low latency. It should be noted that the data generated by IoT devices are generally small in size and therefore, storing them on devices with limited storage is feasible. Furthermore, in many scenarios, such as parking space, contents are often location-dependent, and most of the cached contents are used by users who are in the same location where the content was generated.

To take advantage of caching to the full extent and for better management of cache storage, we propose to group sensors into clusters and then the most powerful node in each cluster selected as the cluster head. The cluster head is responsible for handling the connection with other clusters and can also act as the cache server to coordinate caching decisions inside the cluster. It can distribute cached contents among different devices based

on their energy level, storage capacity, and several other criteria. This can result in better utilization of devices' resources inside a cluster. However, for the overall utilization of network resources, we consider another entity, namely the global SDN/Cache controller (GSCC), that is responsible for orchestrating caching decisions between different clusters and selecting a proper cluster for caching the content. For this purpose, GSCC considers various criteria, including the accumulated resources in the cluster and the location of the cluster in the network graph.

In the proposed architecture, when users ask for content produced by IoT devices in a specific location, instead of heading to the cloud server, we can use GSCC knowledge to identify the cached content's location and forward the request to the cluster storing the content. As mentioned, using CCN in-network caching offers many benefits. However, CCN in its default form was not designed for the IoT. Therefore, to meet the unique characteristics of the IoT environment, some modifications have to be made to the CCN caching architecture. So far, to the best of our knowledge, the related works in the literature do not consider all aspects of IoT, related to caching decision processes, such as redundancy, high bit rate, and unavailability of data when the original producer or cloud cannot be accessed. Therefore, a comprehensive approach for tuning the caching decisions for IoT is necessary, and it is the main objective of this research. In this research, we propose a novel caching mechanism "Extended Multi-Criteria Cooperative Caching" (EM3C) scheme. In EM3C, by considering multiple IoT-related attributes and specifying their relative importance (using methods such as AHP and Shannon's entropy methods), it is decided whether content should be cache or not and also the most appropriate node(s) to store the content are selected based on their available resources. The results of extensive simulations show the efficiency of EM3C in achieving higher cache hit rate and low data retrieval time.

The rest of this chapter is organized as follows: In Section 3.2, we outline research gap and our contribution. Section 3.3 highlights the main challenges regarding the IoT and CCN. We also discuss the proposed system model in this section. In Sections 3.4 and 3.5 we explain our proposed approaches for inter-cluster and intra-cluster caching respectively. Simulation results are presented in Section 3.6. And finally, the chapter is concluded in Section 3.7.

3.2 Research Motivation and Contribution

Building upon the groundwork laid in Section 2.5.1, this research delves into uncharted territories within the domain of caching in IoT networks. The comparative analysis previously conducted reveals a significant gap in existing caching methodologies, particularly, their preference for node-centric strategies that neglect the broader network context and the omits potential of inter-node collaboration. This narrow focus engenders redundancy, with identical content being cached multiple times across the network, and results in cache placements that fail to capitalize on the network's structural intricacies and the diverse capabilities of its nodes. Moreover, the lack of sophisticated routing mechanisms exacerbates the challenge in scenarios rich with routing possibilities, thus reducing cache hit probabilities. A paramount need has been identified for a holistic approach that synergizes content characteristics with nodal capabilities, informed by a thorough understanding of the network's topology, the transient nature of IoT data, and a strategy that orchestrates caching decisions for a network-wide optimum, including smart routing that fosters node cooperation for efficient content retrieval.

Contributions

In response to these challenges, this research posits a novel caching framework that

markedly enhances the efficiency of cache management in IoT networks. Our contributions are manifold and innovatively address the highlighted research gaps:

1. **Multi-Criteria Decision Making for Caching:** We introduce an 'Extended Multi-Criteria Cooperative Caching' (EM3C) scheme, which incorporates a multi-faceted decision-making process. This approach evaluates what content should be cached, and more crucially, determines the optimal cluster and the specific node within that cluster for caching. By leveraging methodologies like the Analytic Hierarchy Process (AHP) and Shannon's entropy, EM3C meticulously considers multiple attributes related to IoT operations, thus ensuring a holistic and informed caching strategy that is attuned to the network's dynamic nature.
2. **Smart Routing via SDN Deployment:** Our framework integrates Software-Defined Networking (SDN) for intelligent routing decisions, enabling a more fluid and responsive network architecture. This not only facilitates efficient content retrieval through cooperative inter-node communication but also optimizes network resource utilization and reduces latency.
3. **Hierarchical Caching Approach:** By adopting a hierarchical structure for content caching, we address the challenge of load balancing and the quest for a global optimum. This structure allows for the distribution of caching responsibilities across different network layers, thereby prolonging node lifetimes and enhancing the overall network sustainability.
4. **Cluster-Based Cache Management:** Central to our proposed framework is the concept of cluster-based cache management, where sensors are grouped into clusters with a designated cluster head for local cache coordination. This strategy is complemented by a global SDN/Cache controller (GSCC) for inter-cluster caching decisions, ensuring an efficient and cohesive caching strategy across the entire network.

These contributions represent a significant leap forward in the field of IoT caching, providing a robust framework that not only mitigates the limitations identified in existing strategies but also introduces a scalable, efficient, and cooperative caching model designed for the complex landscapes of IoT networks.

3.3 System Model

3.3.1. Challenges

Establishing an IoT network can bring a lot of benefits and opportunities, albeit the realization of this idea faces several challenges. The total traffic load produced by billions of devices would be significant and difficult to handle. Furthermore, relying solely on clouds for storing the IoT data would be very costly and results in a long retrieval time. Therefore, considering these challenges, new networking paradigms such as CCN regarded as the key enabler for IoT by many researchers. By exploiting CCN caching capabilities to select and cache the proper contents on intermediate nodes, data can be served with lower latency for upcoming requests. Because of intermittent connections and the existence of power constraint devices in IoT scenarios, caching can also improve the reliability and availability of data. However, the default caching scheme in CCN networks does not entirely fit into IoT scenarios. In a nutshell, the default caching scheme lacks the required intelligence to make smart caching decisions to help prolong the devices' lifetime and also dealt with the probability of original content producer unavailability in the future. The main features of CCN caching are:

- **On-path caching.** Every chunk of content is stored on all the intermediate nodes. In other words, a node must store the content if the content passes through it. Even if these intermediate nodes do not have adequate free storage, they have to make some room for this new content chunk. On the other hand, other nodes that are not located in the content's path have no chance to cache the content, even if they have plenty of free storage space.
- **On-path cache hit.** Content requests can only benefit from the cached content if and only if one of the intermediate nodes on the interest packet's path possesses the cached content. Thus, the probability of a cache hit also depends on the location of the requester and producer of the content.
- **Disunity in path selection and caching decisions.** Each node decides independently whether to cache the content or not. In fact, the nodes do not have any information about the status and capability of the surrounding nodes.

Aforementioned characteristics of CCN default caching scheme leads to several inefficiencies, especially in IoT, including the followings:

- **High level of content redundancy:** Content chunks replicated on each hop, the "on-path caching" feature, results in a huge amount of redundant contents. As a result, the odds of a successful cache hit are minimal, especially in IoT environments in which interest packets may traverse different paths. Furthermore, given the high redundancy and limited storage size of nodes, eventually, the node would not have enough storage to cache new content. In this case, it has to exchange the new content chunk with one of the previous ones.
- **Unbalanced utilization of resources:** Depending on the location of nodes, they can be under heavy load and experiencing high exchange rate and energy consumption, and others may be underutilized.
- **Cached fragmentation:** Every content consists of several chunks. The decision about caching each content's chunk and also the path it should take made separately in each node, so, distinct chunks of the same content may be cached on different nodes [187].

Thus, to deal with the challenges mentioned above, we propose EM3C that uses a novel approach for caching the content in IoT scenarios and is explained in detail in the next section.

3.3.2. Proposed Model

One of the main bottlenecks in CCN caching performance is the disunity in cache management, which results in many inefficiencies, including a large amount of redundant content, suboptimal content selection for caching and inefficient resource management. The main reason for these drawbacks is related to the difference in perception of the optimal conditions from a local and global perspective. To mitigate this problem, we propose a four-level hierarchical model, as depicted in Figure 3.1.

- **Application layer:** In layer 4 of the proposed hierarchical model known as the Application layer, Servers and remote users are located. Servers can be used for long-term storage of data, which is necessary for long-term analyses and assessment of the system.
- **Translation layer:** Layer 3 hosts gateways. Gateways are specific nodes that act as an interface that connects the communication layer to the application layer. A

gateway is responsible for different tasks such as protocol conversion and implementing security policies.

- **Communication layer:** The SDN/Cache controller is located in the communication layer and is used to select the best routes for contents and requests. Considering the global view of the SDN/Cache controller, it is possible to do cache-aware routing, which eliminates the “on-path cache hit” problem of the default caching

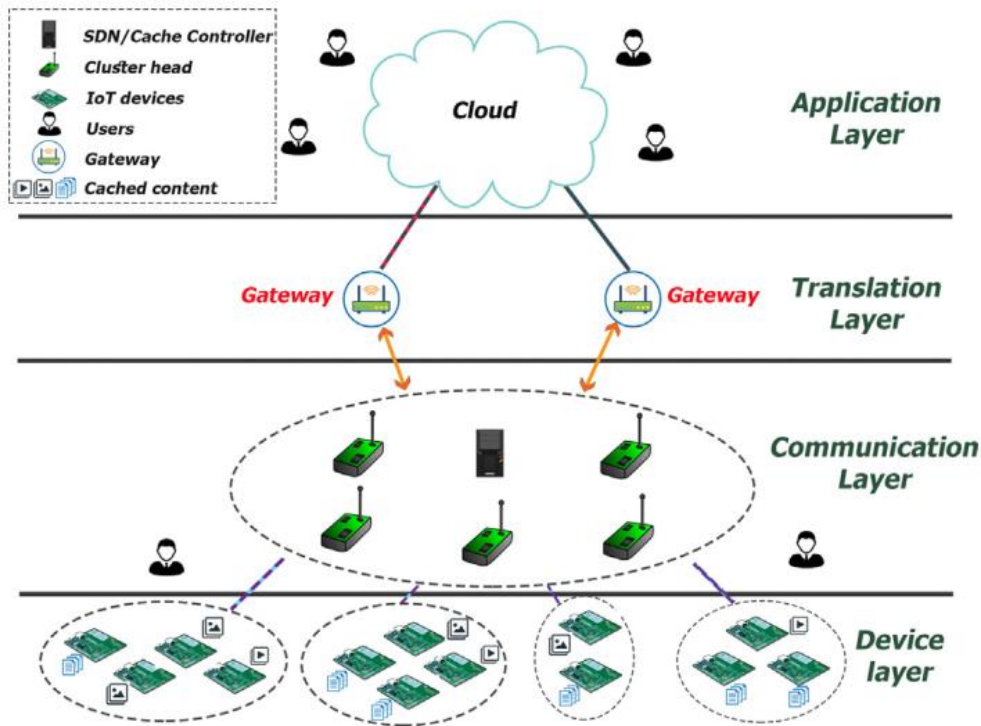


Figure 3.1. Four tier system model for caching in IoT.

scheme. In CCN’s default caching method, the possibility to benefit from a cached content not only depends on the location of the cached content but also the location of the user. In other words, the cached content can only be exploited if it is located on the path between the user and the source. By intelligently forwarding requests toward the cached content using SDN/Cache controller capabilities, we can use the cached content in the vicinity instead of heading to the server located far away. The SDN/Cache controller also acts as a global cache controller and is responsible for making global decisions about the content caching. This procedure consists of selecting a cluster for storing the content and notifying the cluster head about this decision.

Figure 3.2. describes the structure and elements of the communication layer. In the proposed architecture, nodes are grouped in several clusters, and in each cluster, a node is selected as the cluster head. The cluster head is responsible for establishing communication with other clusters and also acts as the local cache controller. They can forward requests to the appropriate destination based on flow entries received by SDN/Cache controller. SDN controller can use the OpenFlow protocol to communicate with the cluster heads. Cluster heads also act as a local cache controller and are responsible for coordinating caching decisions inside the cluster, including whether or not to accept the SDN/Global cache controller’s offers about

caching a content. Another duty of the local cache controller is to choose the proper node(s) based on their capabilities for storing the cached content. Regarding the higher processing power of cluster heads, they are responsible for analyzing cluster member capabilities and assign caching tasks accordingly.

Cluster heads are also responsible for data aggregation. An important aspect of IoT applications in many scenarios, including parking space statuses is that for long-term storage and analysis, IoT applications, mainly interested in a single aggregated data that summarizes the collected data over a period of time such as 15 min or an hour. An example can be the percentage of time a specific parking slot is empty during rush hours, or in the case of surveillance cameras, we can aggregate observations in which data can only be cached if an anomaly detected. This property of IoT data can also be used to decrease the number of unnecessary transmissions and reduce power consumption.

The last entity in the communication layer is local users. These users are located at the same site with IoT devices, and they can communicate directly with the IoT network. Local users, as other users, are the consumers of contents produced by IoT networks. However, because of their location, it would be excessively beneficial if they can access the cached content located in their proximity instead of heading to the server. This results in better QoS and alleviates excessive access to the cloud server.

- **Device layer:** Layer 1 or device layer is where Sensors and in general IoT devices located, which produce and transmit contents regularly. These devices are equipped with storage and are capable of caching the content for later uses.

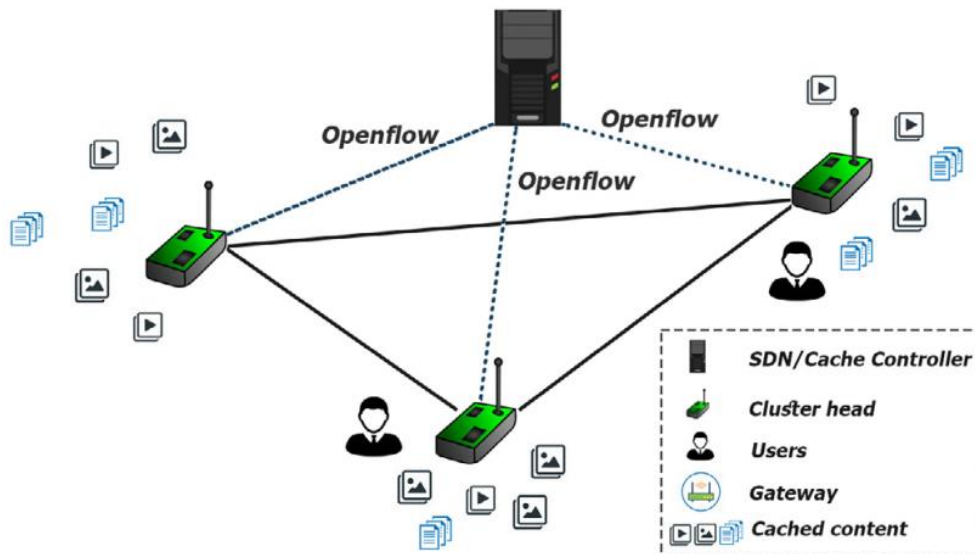


Figure 3.2. Communication layer architecture and elements.

3.4 Inter-cluster global decision making

Upon producing content, the cluster head will be notified, and in turn, the cluster head informs SDN/Cache controller. The SDN/Cache controller selects a gateway and determines the best route to transfer the content to the cloud, based on the overall delay. Delay is calculated by considering the propagation delay and queue length in the gateway. SDN/Cache

controller also selects a proper cluster for caching the content based on its global view of the network and sets the appropriate entries in the flow tables to point to this cluster. It is desirable if the cluster selected for caching the content is located on the path to the selected gateway. This prevents an increase in delay and reduces the number of clusters involved in content transmission. For this purpose, as shown in Figure 3.3, the SDN/Cache controller creates an STP rooted in the source cluster and only selects clusters along the path from the root to the selected gateway. The only exception is that none of the clusters located on the path can cache the content.

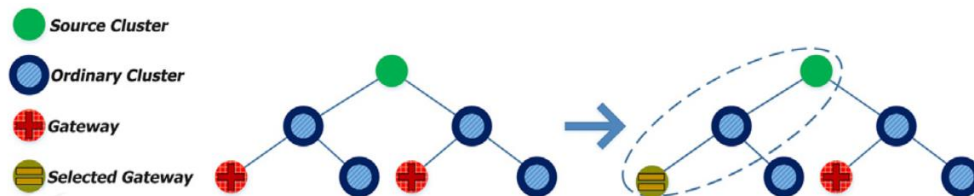


Figure 3.3. Four tier system model for caching in IoT.

To select the most suitable cluster for caching the content, the SDN/Cache controller considers several criteria as follows:

- **Accumulated storage:** accumulated storage is the sum of the free storage capacity of all nodes in the cluster.
- **Accumulated energy:** the overall remaining energy of nodes in the cluster is another metric used by the SDN/Cache controller to avoid early failures.
- **Distance from the source cluster:** because of locality characteristics of most IoT data and requests in many IoT applications, it is preferable to cache the content on the source cluster or close to that. Therefore, the SDN/Cache controller considers the distance from the source cluster as a metric to select the most suitable cluster for caching the content.
- **PLSI:** The characteristics mentioned above mainly focused on a single cluster capability to cache content. However, for a truly efficient caching scheme, we also need to consider the connections and the position of each cluster in comparison to other clusters. To this end, we consider the IoT network as a graph where each cluster is represented as a node. Then we calculate the betweenness for each node. Vertex betweenness is an indicator of the degree of centrality for nodes in a graph. To calculate betweenness centrality, consider the shortest path between every pair of clusters and count how many times a cluster can interrupt the shortest paths. Consequently, clusters with higher betweenness are the clusters located on the center of the network acting as a hub for most of the communication, and caching content on these central clusters will result in shorter average retrieval time and lower hop counts to get the requested content. However, because of the limited storage size, it is impossible to cache all contents on the clusters with higher betweenness also, considering the central cluster's role as a key communications point, extreme use of these clusters for caching, which results in an early power outage can cause many problems for other cluster's communications. Hence, it is crucial to establish some mechanisms to reduce the load on central clusters. For this purpose, we use central clusters only for caching popular content that is in high demand and assigns unpopular content to edge clusters. By doing so, we make sure

that we use the central cluster's sources just for high-value content. Thus, different situations regarding the popularity of the content and the centrality of a cluster arise. These situations are summarized in Table 3.1.

Popularity of content	Centrality of cluster
High	High
High	Low
Low	High
Low	Low

Table 3.1. Possible scenarios in PLSI calculation.

It is also important to pay attention to the concept of popularity in IoT networks. In IoT, contents are transient and are only valuable for a certain amount of time known as Expiry date, so it is not useful to calculate individual content's popularity. Therefore, based on the location-based nature of many IoT applications, we consider popular clusters instead of popular contents. Contents produced by popular clusters are more likely to be requested by users; however, the popularity of a cluster can change over time. For instance, take the scenario of smart parking space management. In the middle of the week, clusters located around government offices may be considered as popular clusters; however, on the weekends, clusters located around shopping centers and entertainment centers would be considered as popular. To specify the interplay between the popularity of content and centrality of a cluster, we introduce a new metric, namely popularity and location-based suitability indicator (PLSI), which is calculated as follows:

$$PLSI = \frac{N - |p - b|}{N} \quad (3.1)$$

" N " is the number of clusters. " b " is betweenness rank of the cluster examined for caching the content, and p is the popularity rank of the cluster that generated the content. The $PLSI$ will have a value close to one when the popularity rank of the content and betweenness rank of the cluster are both high or low; otherwise, it will have a value close to zero.

Based on the metrics mentioned above, the SDN controller selects a cluster for caching the content. However, in the presence of different criteria, it is important to specify each metric's relative importance in selecting the most suitable cluster. Therefore, we use Shannon's entropy method [188] for dynamically calculating each metric's weight. For this purpose, we use the following equations:

$$k = \frac{1}{\ln(m)} \quad (3.2)$$

m is the number of clusters; in the next step, we calculate the divergence of each criterion.

$$E_j = -k \left(\sum_{i=1}^m a_{ij} \cdot \ln(a_{ij}) \right) \quad (3.3)$$

$$D_j = 1 - E_j \quad (3.4)$$

a_{ij} is the normalized value of criteria "j" in the cluster. Finally, we calculate each metrics weight as follow:

$$W_j = \frac{D_j}{\sum_{j=1}^n D_j} \quad (3.5)$$

After calculating each metric's weight, SDN/Cache controller should rank clusters based on their suitability. However, considering different criteria with non-homogeneous scales, selecting appropriate clusters could be challenging. Therefore, we use a modified version of the multicriteria decision-making method called TOPSIS, which explained in the following section.

SDN controller creates a table consisting of clusters (indicating by Cl_1 – Cl_n) as rows and different criteria (m_1 – m_4) as columns and intersection of each column and row which can be represented by a_{ij} indicates the value of the associated metric for the cluster.

	m_1	m_2	m_3	m_4
Cl_1	$a(1,1)$	$a(1,2)$	$a(1,3)$	$a(1,4)$
Cl_2	$a(2,1)$	$a(2,2)$	$a(2,3)$	$a(2,4)$
\vdots	\vdots	\vdots	\vdots	\vdots
Cl_n	$a(n,1)$	$a(n,2)$	$a(n,3)$	$a(n,4)$

Regarding the different scale of m_i , we need to normalize these values; however, instead of using the default normalization method of TOPSIS we use the following equations to normalize values:

$$b_i^+ = \frac{\text{Max } a_{ij} - \text{Min } a_{ij}}{\text{Max } a_{ij}} \quad (3.6)$$

$$b_i^- = \frac{\text{Min } a_{ij}}{\text{Max } a_{ij}} \quad (3.7)$$

$$n_{ij} = b_i^- \left[\frac{((a_{ij} - \text{Min } a_{ij})(b_i^+ - b_i^-))}{(\text{Max } a_{ij} - \text{Min } a_{ij})} \right] \quad (3.8)$$

where b_i^+ and b_i^- are upper and lower bound of normalized values accordingly, which represent the best-case and worst-case scenarios, also n_{ij} is the normalized value of metric i for cluster " j ".

It is worth mentioning that the above-mentioned formulas for normalization belong to positive criteria, and for normalization of negative criteria, we have to make some modifications in these formulas. For positive criteria, such as accumulated storage and the overall remaining power, the more they are, the better they are, but in contrast, for negative criteria such as distance from the source cluster, it is desirable to have lower values.

In the next step, we multiply the Normalized matrix in the weight's matrix obtained by the Shannon entropy method in Eq. (3.5).

$$t_{ij} = n_{ij} * W_j \quad (3.9)$$

In the next step for every cluster " j ", the average difference between the value of each criterion for the current cluster and best and worst-case scenarios are calculated by using the following equations:

$$D_j^+ = \sqrt{(t_{ij} - b_i^+)^2}, \text{ where } i = 1 \dots 4, j = 1 \dots n \quad (3.10)$$

$$D_j^- = \sqrt{(t_{ij} - b_i^-)^2}, \text{ where } i = 1 \dots 4, j = 1 \dots n \quad (3.11)$$

Later, we calculate the relative closeness of each criterion to the ideal solution using the following equation:

$$R_j = \frac{D_j^-}{D_j^- + D_j^+}, \text{ where } j = 1 \dots n \text{ and } 0 \leq R_j \leq 1 \quad (3.12)$$

The closer the value of R_j is to one, the better the cluster is. If R_j is equal to one, it means that the cluster " j " has the best situation in all criteria.

Finally, the cluster with the highest R_j selected as the potential cluster for caching the content and SDN/Cache controller will inform the cluster head about this decision. As will be explained in the next section, the cluster head can either accept or reject this offer.

3.5 Intra-cluster local decision machining

The cluster head in the selected cluster should decide whether to accept the SDN/Cache controller offer and cache the content or reject it, to do this the cluster head should determine the value of suggested content and compare this value with the values of previously cached contents. Another task of the cluster head is selecting the suitable node or nodes for caching the content. Thus, we establish two different processes, one for selecting candidate nodes to store the content and another for determining the value of offered content. In the following sections, we discuss these two processes. Algorithm 1 describes the tasks done by a cluster head on receiving new content. Each received content contains information about its size,

time of production, and a field that indicates if the content is suggested by the SDN/Cache controller for caching on the cluster. This information will be used by the cluster head for handling the content.

Algorithm 1: performed by a cluster head on receiving a new content

```

1. Function content_recieved(content, free_storage) do
2.   if (content→suggested_by_SDN_for_caching()) do
3.     candidate_nodes = determine_candidate_nodes();
4.     content_value = calculate_content_value(content);
5.     if (free_storage >= content→required_storage) do
6.       cache_the_content(content, candidate_nodes);
7.     else
8.       cache_replacement(content, content_value);
9.     end
10.  end
11.  forward_content_to_next_cluster_on_the_path(content)
12. end

```

3.5.1. Node selection

After receiving an offer from the SDN/Cache controller, if this offer is accepted, the cluster head has to choose candidate nodes. Candidate nodes are a group of nodes selected by the cluster head to cache the content. By choosing appropriate nodes for caching content, we can improve the stability and overall lifetime of the network. Regarding the nature of IoT devices, we choose several attributes as follows:

1. **Free storage space:** to be able to cache the content, having enough free storage is one of the main prerequisites. Higher free storage is appreciated, but even if the free storage is not enough for caching the content, by leveraging the proposed replacement policy, we can exchange new content with previously cached contents. However, very low free storage is a sign that if the node is selected to store the content, a lot of stored chunks will need to be removed. Thus, to elevate the efficiency of the algorithm, we ignore the nodes in the node selection process if their free storage is lower than a threshold value.
2. **Energy level:** the power level is another crucial metric, which represents the remaining energy level of the node. Energy is the node's most valuable asset which should be consumed very carefully to prevent early failures. To ensure a long lifetime of nodes, we ignore the node in the node selection process if the energy level drops lower than the threshold value.
3. **Distance from cluster head:** another critical metric is the distance between the node and the cluster head. Every communication with the outside of the cluster should be done through the cluster head. Thus, for serving the request with the cached content, this content should be sent to the cluster head, and the cluster head then forward the content toward the user. However, nodes located in the vicinity of the cluster head consume less energy to communicate with the cluster head, but with increasing the distance between the node and the cluster head, the

node has to consume more energy to communicate with the cluster head. Therefore, as shown in Figure 3.4., we consider clusters as multiple concentric circles, and at the center of these circles is the cluster head. To reduce energy consumption, we prefer nodes located in circles close to the center because they consume less energy to communicate with the cluster head. To specify the circle in which a node is located, a beacon is sent by the cluster head to each node and by calculation of the round-trip time, the circle in which the node "j" belongs, is calculated as follows:

$$C_j = \left\lceil \frac{RT_j}{\text{Min}(RT_i)} \right\rceil, \text{ where } j = 1 \dots n, i = 1 \dots n \quad (3.13)$$

where RT_j is the round-trip time of the beacon between node "j" and the cluster head, and C_j is the circle in which the node j located

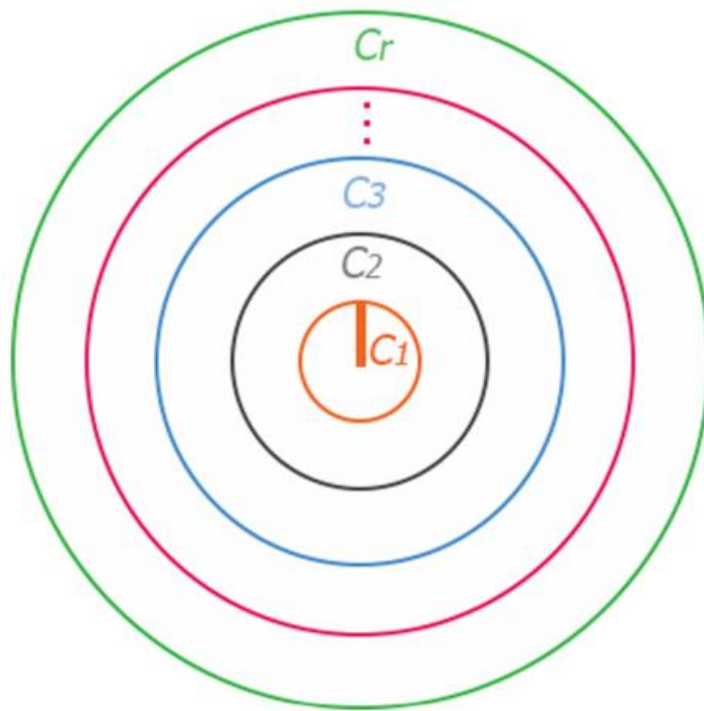


Figure 3.4. Four tier system model for caching in IoT.

- **Number of similar sensors:** the last metric used by the cluster head to select nodes is the number of similar nodes. It is very common in IoT to have several similar sensors in an area, mainly for redundancy purposes. Regarding this fact, it is essential to consider the number of similar nodes in each cluster. For instance, if there is only one node from one specific type, and we lose that node because of the power outage, then there will not be any alternatives that can produce similar data but on the other hand, if there are several sensors from one type, even if we lose one of them, there are other nodes that can produce the same type of data.

The integration of the above-mentioned metrics yields a single value, which is called caching weight. To calculate the caching weight, the significance of each of the aforementioned metrics should be determined. Considering the limited processing

higher values extend the lifetime of nodes and lower values improve the cache hit rate. Figure 3.5, illustrates the percentage of devices that failed (run out of energy) after 100 days, using different α rates. As can be seen, higher α results in lower failure rates thus it is reasonable to choose a higher α .

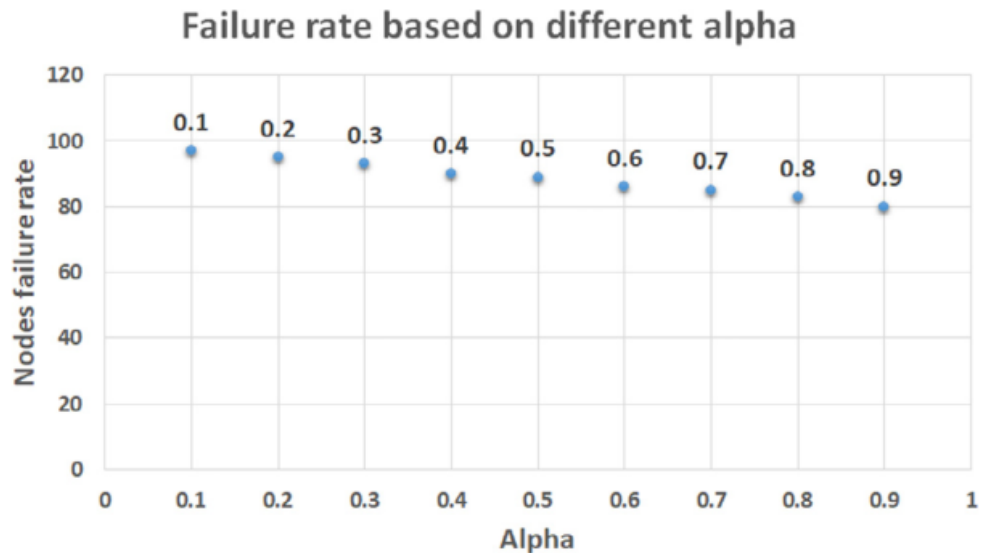


Figure 3.5. Failure rates for different alpha values.

However, as depicted in Figure 3.6, lower α usually results in better cache hit rate, especially in the first 60 days. Therefore, there should be a trade-off that provides reasonable cache hit rate and failure rate. One can set the value of α according to the scenario in-hand. Without loss of generality and according to our simulations results, we have taken α equal to 0.7 in the performance evaluation of this chapter.

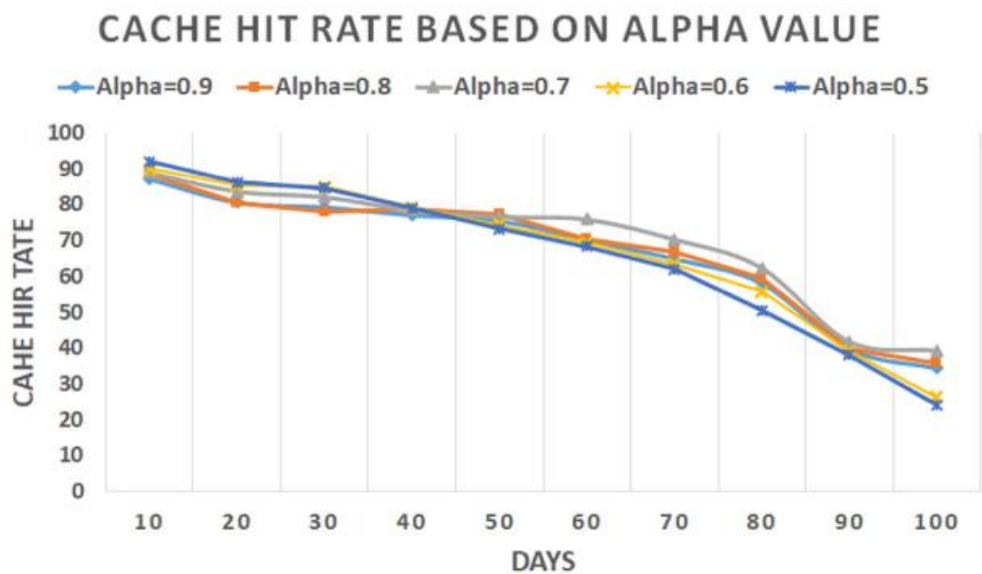


Figure 3.6. Cache hit rates for different alpha values.

3.5.2. Determining candidate contents

The main objective of this process is to select contents for caching so that the benefits of caching (reducing access time, increasing accessibility, etc.) are more significant than the costs of caching (storage consumption, energy consumption, etc.). Thus, by considering the characteristics of contents in the IoT environment, we select three criteria for evaluating the value of the contents.

4. **Freshness:** one of the significant aspects of contents produced in the IoT environment is their short lifetime. The transient nature of the contents of IoT means that content is regarded as invalid after a certain period, which defined by the content producer. Furthermore, the requests in this environment are usually for recently generated contents. To evaluate the freshness of content, we use the following equation:

$$F_i = T_i - Age_i \quad (3.17)$$

where T_i donates the lifetime of the content i that is the expiry date specified by the producer and Age_i is the period between the content's time of production and the current time. Content is assumed to be invalid if its age is more than this value, and it can no longer be selected for caching in this case.

5. **Popularity:** popularity is another important metric that should be taken into account in cache selection policies. Higher popularity means a higher chance of being requested by users. However, there is a fundamental difference between the popularity of content in the IoT environment and the popularity of regular content. As mentioned previously, contents in IoT are transient with a limited lifetime. Furthermore, consumers of IoT content usually are interested in location-based data, for instance, the temperature in a specific location or free parking space in a neighborhood. Thus, it is reasonable to associate popularity with clusters and not individual contents.
6. **Content size:** the last content related metric is the content size. As discussed previously, the caching decision is applied to all chunks of the content. As a result, the content is either cached as a whole or put away entirely. Thus, it is vital to consider the number of forming chunks of the content. Caching large content which is requested frequently can save a lot of bandwidth and energy, but at the same time, large contents can take a lot of storage space and prevent other useful contents from caching; therefore, it is essential to preserve a good balance.

The integration of the above-mentioned metrics results in one single value, which is called caching weight. However, as in the candidate node selection process, these metrics values needed to be normalized. In the next phase, the relative importance of each metric should be determined. For this purpose, as in the selecting candidate nodes process, we use the AHP. After using the AHP to calculate the weight of each criterion, we calculate the content's value as follows:

$$W_i = \sum AW_j \times r_{ij} \quad (3.18)$$

where AW_j is the weight of metric "j" calculated by AHP. r_{ij} is the normalized value of content "i" in metric "j", and W_i is the value of content "i".

Then we use calculated content's value to decide about whether or not the content should

be cached. For this purpose, we consider candidate nodes, if the sum of free storage on candidate nodes exceeds the offered content size, the offer is accepted, but if not, we have to follow cache replacement policies. To this end, we compare offered content's value to the values of previously cached contents on the candidate nodes. However, due to the distribution of cached content among nodes in the cluster, some portion of the existing content on the candidate nodes may be stored on non-candidate nodes. We show the portion of content stored on candidate nodes as γ and use it to calculate a new weight for the previously cached content as follows:

$$CW_i = \gamma \times W_i \text{ where } 0 < \gamma \leq 1 \quad (3.19)$$

where W_i is the original value of the content, and CW_i is the newly calculated value, which is proportional to the fraction of the content cached in the candidate nodes.

Later, we arrange the contents in descending order based on their CW_i , and starting from the content with lowest CW_i , we select the content and add the value of CW_i 's until the total size of these contents is greater than the space required to cache the offered content. We call these selected contents, potential discarding candidates (*PDS*). Next, the total CW_i of these contents, which is denoted by SCW is compared with the offered content's W_i . If offered content's weight is higher than SCW , then the cluster head decides to discard *PDS* and cache

Algorithm 2: Cache replacement algorithm

```

1. Function Cache_replacement ( $w_i$ ,  $CW[]$ ,  $Sum\_of\_PDS\_storage$ ,  $required\_Storage$ ,  $new\_content$ ) do
2.   while ( $sum\_of\_PDS\_storage < required\_storage$ ) do
3.      $SCW = \sum CW_i$ 
4.   end
5.   if ( $SCW < W_i$ ) do
6.     discard_PDS();
7.     cache_content( $new\_content$ );
8.   else
9.      $should\_consult\_with\_SDN = (rand() < \frac{w_i}{SCW}) ? true : false;$ 
10.    if ( $should\_consult\_with\_SDN$ ) do
11.      Find_next_suitable_cluster( $new\_content$ );
12.    else
13.      Forward_content_to_cloud();
14.    end
15.  end
16. end

```

the new content; otherwise, the cluster head rejects the offered content. However, if the cluster head decides to reject the SDN/Cache controller's offer (which takes place with the probability of W_i/SCW), it may ask SDN/Cache controller to select another cluster for caching the content. The algorithm 2 is responsible for deciding about replacing previously cached content with the new content.

3.6 Performance evaluation

We use Omnet++ simulator to evaluate the performance of our proposed caching scheme. We implement CCN as an Overlay network on top of the IP layer. The simulation results are obtained by averaging over 60 runs. To evaluate the performance of the proposed scheme in

comparison to similar methods, we also implemented two other caching schemes: M3C [17] and MACD [140]. Both M3C and MACD are data caching methods that are specifically optimized for IoT networks, and therefore, they are suitable for comparing with EM3C under similar conditions. A basic simulation scenario consists of 100 CCN nodes, which are grouped in some clusters, and each cluster managed by a cluster head. Nodes are uniformly categorized to 3 types (representing various types of nodes in a real-life IoT scenarios): type 1, type 2 and type 3 with respectively storage of 17 MB, 22 MB and 27 MB. The nodes periodically generate location-dependent information and send it to the cluster head, upon receiving this data; the cluster head consults the SDN/Cache controller and tries to send the received content to the server. However, when the cluster head does not have direct access to the server, the data must pass through several intermediate clusters to reach the server. There may also be multiple paths to the server; in this case, the SDN/Cache controller selects the best path and notifies the cluster heads.

Furthermore, the most appropriate cluster for caching the content is selected by the SDN/Cache controller so that requests for content can be responded through cached information. Content requests are generated by multiple users who move randomly through the simulation environment and periodically issue content requests. However, due to the location-based nature of the generated contents, most requests issued by users are for contents created at nearby clusters. The distribution of requests also follows the Zipf distribution, with about 80% of requests being for newly produced contents and about 20% for older contents.

The cache hit rate is one of the most important criteria used to evaluate the performance of the proposed method. The cache hit rate is the percentage of requests that are responded using cached content without having to interact with the server. As a result, a method with a higher cache hit rate can significantly help to address the problems of the IoT environment, including the availability of content at any time. As can be seen in Figure 3.7, EM3C offers a higher caching rate than all other methods. Furthermore, over time, as IoT devices lose their ability to cache data, all methods are experiencing performance degradation, but the EM3C's performance loss rate is quite gentle, and even after 100 days, we see its acceptable performance.

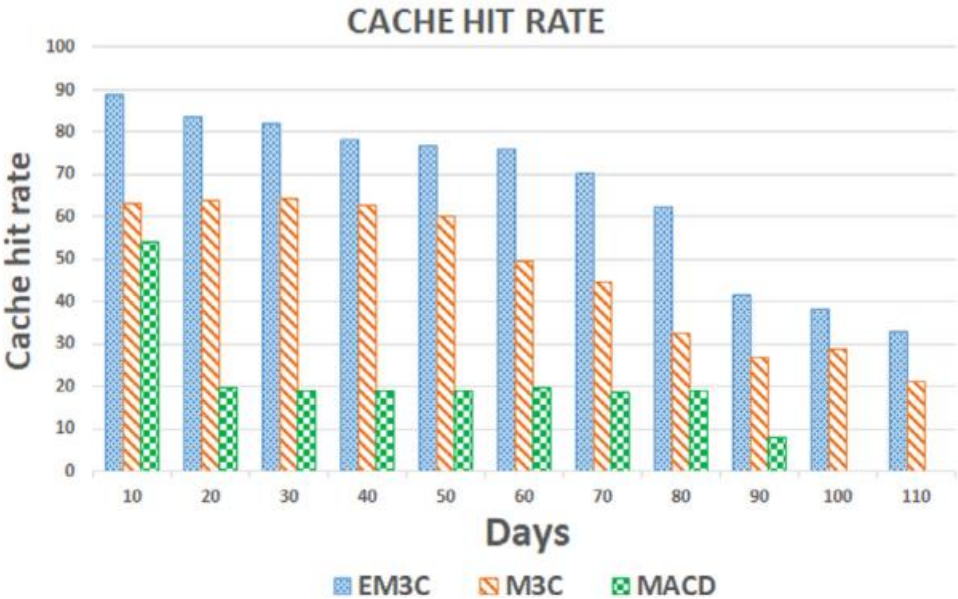


Figure 3.7. Cache hit rate comparison of EM3C, M3C and MACD.

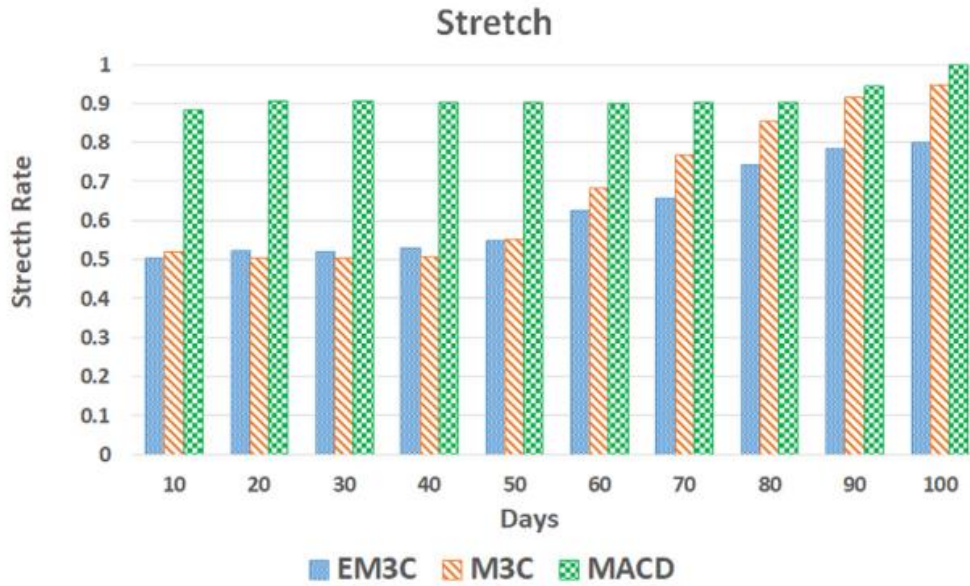


Figure 3.8. Stretch comparison of EM3C, M3C and MACD.

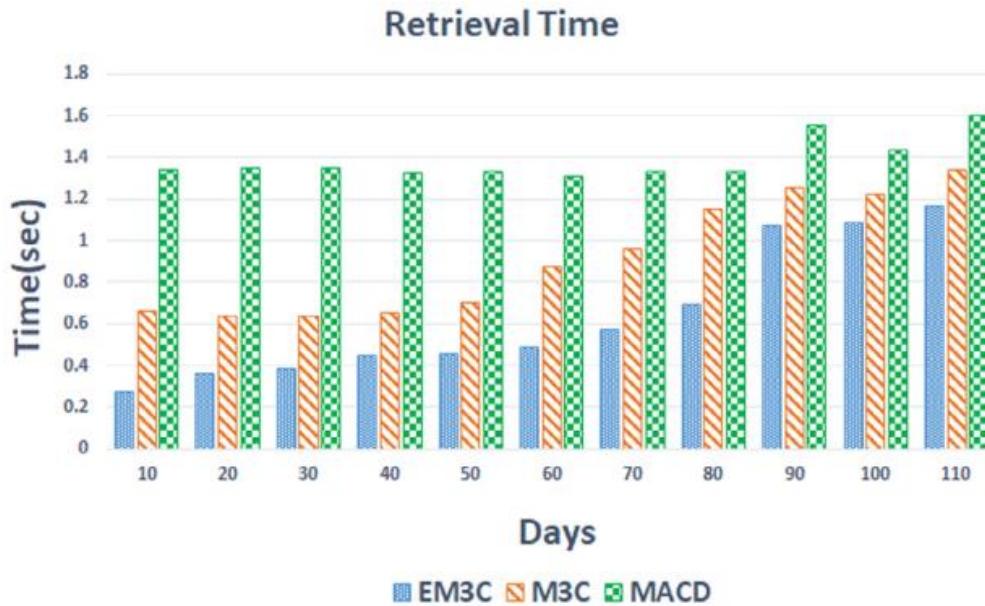


Figure 3.9. Retrieval delay comparison of EM3C, M3C and MACD.

Since more hop counts mean passing the request through more clusters and involvement of more nodes, this, in turn, results in more processing, more energy consumption, and more delay. Thus, it is essential to reduce the average hop count. So, we considered stretch factor [190], which is defined as the number of hops that the request has traveled before reaching to the content divided by the number of hops from the user and cloud server. Stretch factor is a number between 0–1 where lower values are more desirable. Figure 3.8 shows that the stretch factor in EM3C is higher than the other two methods, because, by efficient use of available nodes' storage, EM3C can achieve higher cache hit, which enables it to provide the requested content by using cached data in adjacent clusters without heading to the server, so having a lower stretch factor. However, as time passes, the energy level of nodes decreases, and thus, the number of contents being cached is reduced. This is because, in the proposed caching schemes, the nodes' energy level contributes to the ability of nodes for caching

content. This, in turn, leads to sending more users' requests to the cloud resulting in a higher stretch.

One of the main factors in investigating the efficiency of the caching mechanism is the retrieval delay, which is the time between requesting the content and getting the content. For delay-sensitive apps, this can be very important to keep retrieval time below a specific threshold. Furthermore, the lower delay can improve the quality of experience considerably, and as illustrated in Figure 3.9, retrieval delay in EM3C is always lower than MACD and M3C. This can be explained by the intelligent routing of requests by SDN toward the cluster possessing the content. However, as explained earlier, less content is cached over time, and more requests are sent to the cloud for the response. Due to the long response time from the cloud compared to the local network, this, in turn, leads to increased latency.

A good caching scheme should avoid redundancy and cache of duplicate content. Redundancy leads to a waste of resources and prevents caching of new content. Diversity expresses the ratio of unique content stored on all nodes in the network. Diversity varies between $[1/N, 1]$, Where "N" represents the number of nodes in the network. If diversity tends to $1/N$, there will be multiple replicas of content in the entire network; otherwise, if diversity tends to 1, there is no redundancy and only distinct chunks can be cached. As can be seen in Figure 3.10, the diversity of EM3C is always equal to 1, which is the highest possible value. EM3C's global cache management prevents any redundancy, and there is only one cache per content across the network. This may raise availability concerns due to existing only one cached version of any content. However, EM3C avoids caching content in nodes that are likely to run out of power soon, so the chances of a node containing cached content to shut down are meager. Besides, contents are also accessible through the cloud even if the node crashes, and if the content is valid, it will be re-cached after the first request is answered

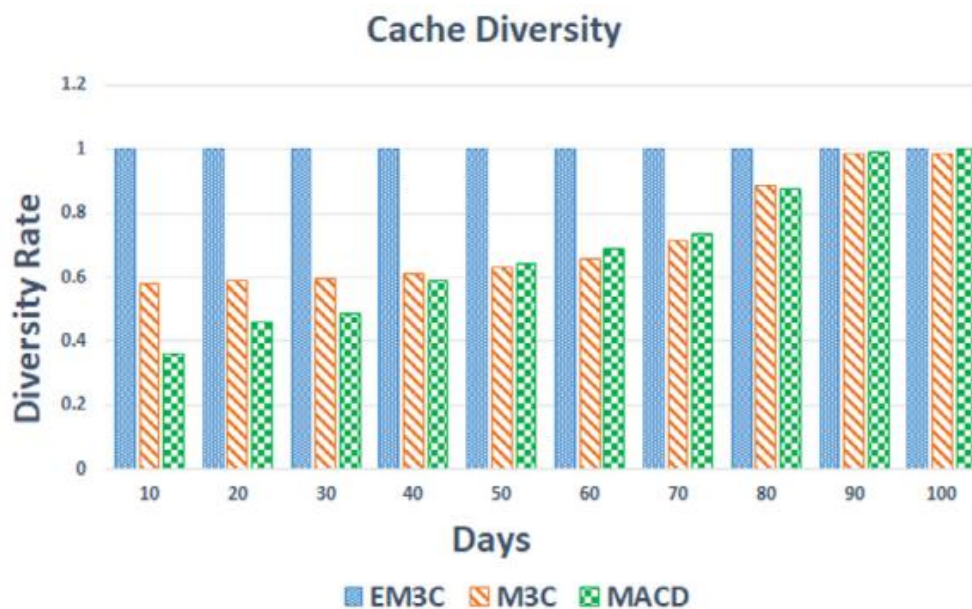


Figure 3.10. Diversity comparison of EM3C, M3C and MACD.

through the cloud. Comparing the other two methods: the M3C initially performs better because of the semi-global decisions made by the cluster-head. However, in MACD, due to inefficient resource management, over time, more nodes experience a power loss. By reducing the number of nodes capable of caching content in MACD, diversity will increase. After 100 days, due to the significant decrease in active nodes, diversity tends to 1 in all methods.

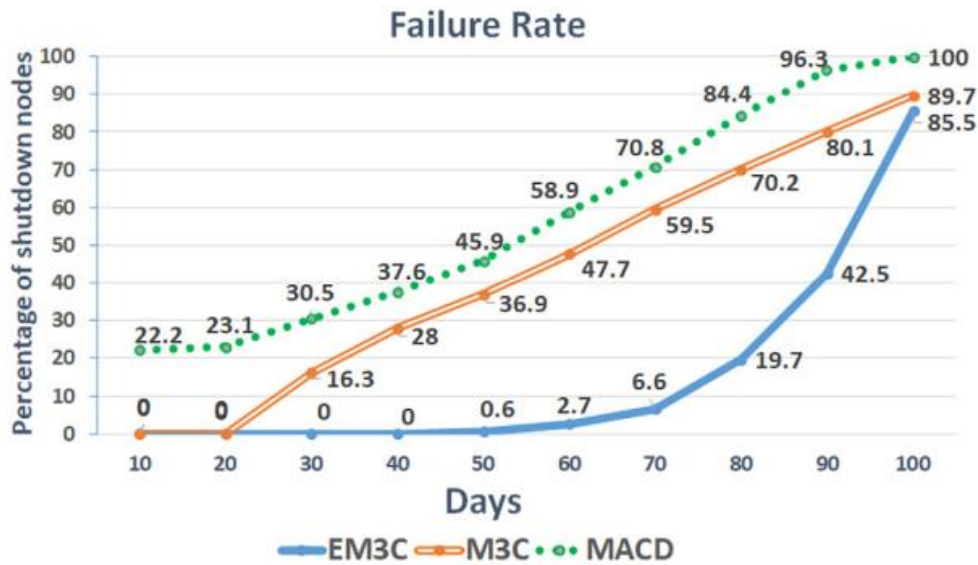


Figure 3.11. Nodes' failure rate comparison of EM3C, M3C and MACD.

Extending the lifetime of devices should be the goal of any proposed scheme for IoT. Ideally, a caching scheme must consider devices' power level in assigning caching tasks to devices and avoid using nodes with a low-power in this procedure. A good caching scheme prevents early node failures by proper management of available resources. To evaluate the efficiency of each caching scheme regarding the nodes' lifetime, we consider the failure rates of the nodes.

As illustrated in Figure 3.11, From the very first days, the MACD method experiences nodes failure. In the MACD method, the nodes with lower capabilities fail in the early days and over time the more powerful nodes fail as well, such that before the 90th day, the entire IoT system crashes. It is worth noting that after a hundred days in all methods, the IoT system has lost most of its nodes due to power outages. However, a caching scheme that can keep most nodes active until the last days of the system lifetime is very desirable. In the EM3C method, up to 60% of nodes are active by day 90 and can sense the environment and send information to the server or users.

To examine the effect of nodes cache size on the performance of the proposed algorithm, we executed the simulation by considering different nodes' storage size. As mentioned before, three types of nodes with different energy levels and storage size are available in the system. Therefore, we increased (decreased) the basic storage scenario by 20% and 40% which results 5 storage scenarios and reported the average storage size in each scenario in the horizontal axis of Figure 3.12. To test the performance in stable conditions, the results are considered in the range of 30th to 70th day, which we call the stability period. As can be seen in Figure 3.12, increasing the amount of node storage size leads to an improvement in the cache hit rate because more information can be stored on the node. On the other hand, fewer nodes need to be involved to cache the content, which in turn improves the performance of the proposed method.

Finally, we study the effect of topology size on the performance of algorithms. In the basic topology size, 100 nodes are considered for EM3C evaluation, which are divided into 8 clusters. Now the simulation is repeated with different numbers of nodes and the results are depicted in Figure 3.13. It should be noted that the presented results belong to the stability period (day 30–70) and also the number of users and therefore the number of requests in

different topologies are remained unchanged. As follows from Figure 3.13, we can see higher cache hits rate in topologies with a higher number of nodes. This is because an increase in the number of nodes, on the one hand, means an increase in available storage capacity and on the other hand means a greater number of nodes available to participate in the caching process.

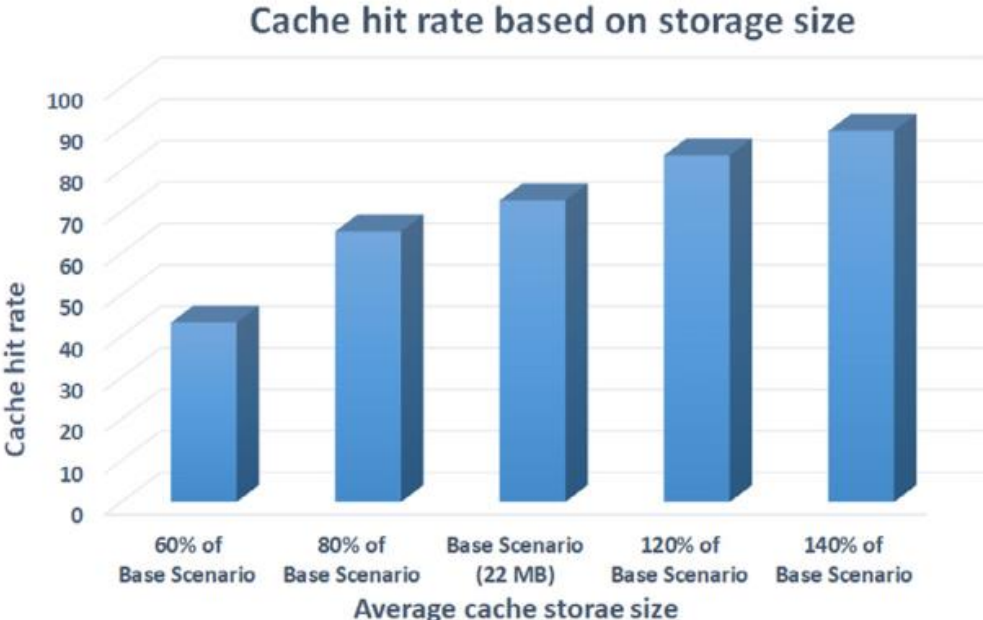


Figure 3.12. Cache hit rate comparison for different storage size.

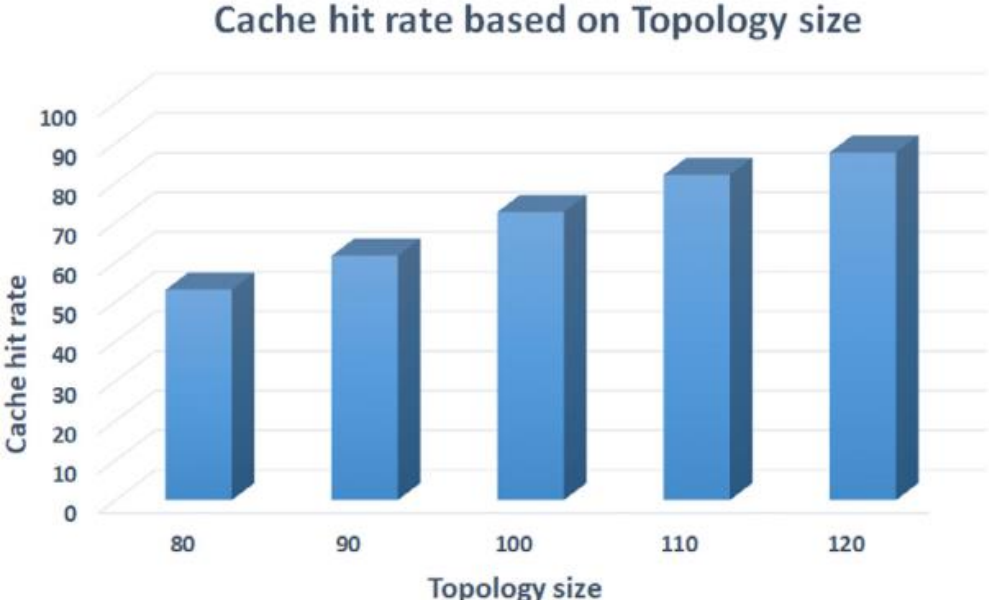


Figure 3.13. Cache hit rate comparison for different topologies.

3.7 Conclusion

In this chapter, a new scheme called EM3C was proposed for data caching in the IoT environment. EM3C increases the lifetime of the IoT devices by optimizing resource consumption and allows for caching more content by reducing the redundancy of cached contents. In the proposed method, by using the SDN routing capability, the requests are intelligently directed to the cached content, resulting in reduced latency and number of hops and increased cache hit rate and network lifetime. SDN also selects the best cluster to cache content using its global view of the network. The cluster head then decides whether or not to cache the offered content, and according to the capabilities of the cluster nodes, distributes the proposed content among the nodes. The results show the superior performance of the proposed approach in improving system performance. The EM3C method provides up to 35% more cache hit rate than other methods. This, in turn, leads to shorter retrieval times, which is up to 60% percent lower than other methods. Furthermore, requests in EM3C take fewer steps to get the content in comparison to other caching schemes. Finally, EM3C can guaranty the standard functionality of the IoT systems until the last days of the system lifetime by preventing early failures.

As next chapter, we intend to consider mobility and try to modify our method to consider nodes movement in the environment. Thus, the decision to cache contents will be made respecting to the speed and direction of nodes.

4. An Intelligent Caching Scheme Considering the Spatio-Temporal Characteristics of Data in Internet of Vehicles

4.1 Introduction

In the past decades, the development of Intelligent Transportation Systems (ITS) has been the focus of governments and organizations in all over the world. The Internet of Vehicles (IoV) is considered an essential component of ITS. As vehicles evolve from simple means of transportation to smart entities with computation and communication capabilities, they have become an integral part of ITS and a smart city. IoV provides an infrastructure for developing applications to provide various services. One of the problems we face with the increasing use of private cars is the high number of accidents and fatalities. In the United States, pre-retirement deaths from car accidents are greater than those from the two leading diseases, cancer and heart disease combined, while nearly half of all deaths among 19-year-olds can be attributed to car accidents [191]. In addition to the loss of life, accidents also impose other harms on people, including emotional suffering, economic costs, and travel time delays for other vehicles. By using vehicle communication facilities in IoV, it is possible to increase the safety of car occupants by sharing emergency information in a timely manner. Also, with access to traffic information, drivers can plan their routes to reduce travel time and fuel consumption.

In addition, as a result of recent advancements in vehicular technology, autonomous vehicles are no longer a distant reality. As a result, passengers will find themselves with a good deal of free time in autonomous vehicles. Passengers can now spend their time to be entertained or work. The interest of these users is mainly focused on multimedia content, including watching and sharing video, audio, and images on different platforms. Delay in receiving such multimedia contents can have a great impact on the user experience of passengers. Additionally, new applications developed for vehicles require the exchange of high-volume data, which puts substantial pressure on the vehicular networks. Edge caching has been proposed as a solution to simultaneously reduce content retrieval latency and backhaul traffic for high-volume data delivery. Consumers only care about the content itself, not where it is retrieved from, and the nature of the network has changed from "where to retrieve data" to "what data to retrieve" [15]. By responding to requests from the cache, the number of requests sent to the backhaul will decrease and users will experience a lower delay. Due to the limited resources of RSUs, it is necessary to develop a mechanism for content placement. In addition, the frequent connection loss and connection re-establishment due to limited range and high mobility, pose significant challenges to resource management in the IoV. Therefore, the caching scheme should also consider the mobility of vehicles.

In addition, the exploration of edge caching strategies within the IoV, as detailed in Section 2.5.2, uncovers a significant reliance on a singular metric for caching decisions: the popularity of content. This predominant focus simplifies the multifaceted spatiotemporal dynamics and the rich diversity of vehicular applications into a mere count of content requests. Such methods, while aiming for dynamism, fall short of capturing the evolving landscape of IoV needs. Notably, some advancements have ventured into privacy-preserving popularity predictions. However, a holistic framework for selecting nodes in the learning process and

implementing policies for high-quality model development is notably absent. Additionally, the prevalent trend towards localized caching decisions overlooks the potential enhancements a globally optimized strategy could bring. This approach misses out on maximizing the synergistic effects of node collaboration, leading to inefficiencies in caching mechanisms. Hence, as emphasized in Section 2.5.2, there exists a critical demand for a comprehensive caching model. This model must account for the intrinsic nature and spatiotemporal characteristics of IoV content, fostering cooperative decision-making that aligns with a vision for global optimization. By addressing these outlined gaps, we pave the way for a more sophisticated, efficient, and user-centric caching framework within IoV systems.

In this work, in order to deal with the above-mentioned challenges, first, we consider the vehicular data in two types: emergency data and infotainment data. These applications have different requirements in terms of the Quality of Services (QoS). For example, applications related to vehicle safety require low latency, while infotainment services require high throughput. Therefore, the characteristics of the content should be investigated and a dedicated caching mechanism should be proposed for each type.

For emergency data, we propose the spatio-temporal characteristics Aware, Emergency Content Caching (STAECC). STAECC introduces the Red Zone for emergency events. Safety-related data is local in nature and only relevant to vehicles in a specific region.

For infotainment data, we propose the Federated learning-based Mobility-aware Collaborative Content Caching (FM3C) scheme. FM3C examines the characteristics of this data type and selects the appropriate contents to cache in each RSU. One of the most important metrics for infotainment content is its popularity, which is an indication of how much demand there is for the content in question. Popularity is estimated using the Long Short-Term Memory (LSTM) model according to the historical information of users' requests. However, as users are often unwilling to share their information, we use federated learning to train the model collaboratively while preserving user privacy.

We have particularly focused on ensuring scalability while addressing the unique challenges associated with the two categories of vehicular data: emergency and infotainment data.

1. For safety content, the processing is localized, managed solely by the RSU present at the accident site. We have developed the STAECC, a lightweight mechanism that uses pre-computed weights determined by Multiple Criteria Decision-Making techniques to make caching decisions. This approach sidesteps the need for resource-intensive machine learning, enhancing the scalability of this method.
2. As for infotainment content, we have utilized Federated Learning to ensure privacy and address scalability issues. Each vehicle employs a LSTM model to train a local model based on its request history. Given the considerable computational power of modern vehicles and the manageable size of individual request history datasets, local model training will not pose significant scalability issues. RSUs are then tasked with aggregating these local models, a process that features linear complexity. It's worth noting that the selection mechanism proposed ensures that irrespective of the number of vehicles within an RSU's coverage area, only a selected group of them is chosen for this process, further enhancing scalability.
3. Lastly, the Fog server plays a critical role in coordinating caching decisions among RSUs. Its role primarily features linear complexity, scaling with the number of RSUs and their storage capacities rather than with the number of vehicles. This design ensures that the system can efficiently manage and scale regardless of the volume of vehicular traffic.

This meticulous design consideration ensures that our proposed solutions, STAECC and FM3C, are highly scalable, thus robustly addressing the ever-evolving and demanding needs of IoV systems.

Finally, we introduced a Fog server that determines the appropriate content to be cached in each RSU based on content characteristics.

For content retrieval, we use Software-defined networking (SDN) [192] to intelligently transfer the request to the cached content, and the RSUs can cooperate with each other to respond to the requests and reduce the delay.

The central contributions of this research are manifold and target the complexity of the caching challenge in IoV environments:

- We categorize vehicular data into two distinct types: emergency and infotainment. Recognizing their differing characteristics and requirements, we propose unique and innovative caching strategies for each, namely FM3C and STAECC.
- Unlike most existing studies that predominantly focus on popularity, we adopt a multi-faceted approach to evaluate content for caching. We use multi-criteria decision-making methods, like the Analytic Hierarchy Process (AHP), to calculate the weight of various metrics. These metrics include not just content popularity but also other content attributes, as well as the storage capacities of Road Side Units (RSUs).
- Further, our caching scheme is cooperative, aimed at minimizing redundancy across the network. This approach ensures a more effective use of the limited storage resources at RSUs and ensures a higher diversity of cached contents.
- We propose a pioneering popularity prediction mechanism, rooted in federated learning, tailored to the unique dynamics of the IoV environment. Our method incorporates a novel model aggregation strategy, taking into account the quality of the model trained by individual vehicles.
- Lastly, we introduce an innovative policy for selecting vehicles to participate in the federated learning process. This strategy ensures effective learning while balancing the computational load among vehicles, enhancing the system's scalability.

The remainder of this chapter is organized as follows. We describe the system model in section 4.2. The proposed caching strategies are introduced in section 4.3, we then we present a method for predicting the popularity of infotainment content in section 4.4. Considering different criteria to evaluate the suitability of a content for caching, we will describe the multi-criteria decision-making method in section 4.5. This will be followed by a description of how Fog server make a global decision about content caching for RSUs in section 4.6. The performance evaluation of the proposed caching schemes presented in section 4.7. Finally, conclusions and our future work are outlined in section 4.8.

4.2 System Model

Figure 4.1 provides a visual representation of our system model, a multi-tier vehicular network in an urban area, comprising three distinct layers: the Edge layer, the Fog layer, and the Cloud. The Cloud layer, being the last resort for content retrieval, is accessed when a user requests content that hasn't been cached in the lower tiers. The Fog layer, the second stratum of our system architecture, includes two essential entities: the SDN Controller and the Orchestrator. The Orchestrator plays a crucial role in global content placement. With its

holistic view of the network and the data it receives from all Roadside Units RSUs, it can ascertain the optimal content to cache, thereby reducing storage space wastage and redundancy. The SDN Controller, another key entity, is primarily employed for content retrieval. The SDN Controller allows network administrators to manage the underlying network infrastructure for network applications and services more efficiently, directing requests to the nearest cached content and simplifying the process.

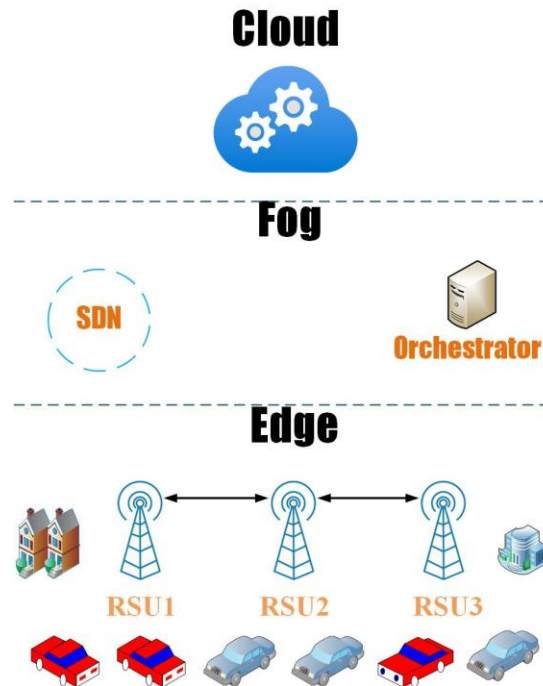


Figure 4.1. The proposed multi-tier vehicular network architecture.

The Edge layer, the final layer of the architecture, consists of a set of RSUs " $R = \{1, \dots, n_r\}$ " and the communication range of RSUs are denoted by D_r . RSUs associated with a Fog server via a reliable link and in turn, the Fog server is connected to the cloud using backhaul links. RSUs and the Fog server are equipped with caching capability with the storage size of each node depicted by " $\{Scap_1, \dots, Scap_n\}$ ". Within the coverage area of each RSU r_i , there are several roads " $S = \{1, \dots, n_s\}$ " and let l_i denotes the length of the road s_i and vm_i is the maximum allowed speed on the road segment s_i . We define a set of contents " $C = \{1, \dots, n_c\}$ " where each content is described by three terms as " $\{z_c, g_c, t_c\}$ ", where z_c is the size of content C , g_c is the category of the content which refers to the topic of the content and t_c is the type of the content, contents are divided into two general types: emergency and entertainment. Emergency contents are usually small in size and although fast delivery is desirable for both types of content, The urgency of the safety contents means that if the content is delivered late, the driver will not have the time to react, and therefore the quick delivery of the safety contents becomes vitally important. Infotainment contents is common and regularly requested by users. This type of content is generally voluminous and quick access to this type of content can have a positive impact on the user experience.

Each RSU is responsible for handling requests received by vehicles within its coverage area. However, due to the limited storage capacity in the RSU, not all contents requested by users are readily available, therefore, the RSU must retrieve the missing contents from other sources such as other RSUs, Fog server, or the Internet. The cached content information (CCI) of an

RSU r_i at time " t " is represented as:

$$CCI_{it} = \{C_{1it}, \dots, C_{nit}\} \quad (4.1)$$

Where $C_{jit} = true$ means content C_j is cached in r_i at time " t ".

The total size of the cached content in any node must not exceed its maximum capacity. Therefore, we have:

$$\sum_{i=1}^{|R|} \sum_{j=1}^{|C|} C_{jit} \times Z_j \leq Scap_i \quad (4.2)$$

To improve the user experience, we develop a method which aims to reduce cost and delivery time by providing content from the nearest neighbor. In particular, delivery time is a function of content size, available link capacity, and distance from the cache storage that holds the content.

Let " $V = \{1, \dots, n_v\}$ " represents a set of vehicles, which enter the city in random periods, each vehicle has a trajectory " $\{S_1, \dots, S_j\}$ ", that can path through the coverage area of several RSUs and v_{ijk} is the j -th moving vehicle on road s_k in the coverage area of RSU r_i . In our system model, vehicles send content requests to the connected RSU and each vehicle records its request history, RSUs also anonymously save the history of requests they have received.

4.2.1. Mobility Model

In our model, there are $|V|$ vehicles in the environment and each vehicle have a trajectory as " $\{S_1, \dots, S_n\}$ ". The speed limit for each section of the road is defined as " $\{vm_1, \dots, vm_n\}$ ". However, it is not realistic to consider the speed of vehicles independently as in many previous research studies. Therefore, in this work, we describe the traffic situation dynamically. For this purpose, we first describe the following macroscopic variables that are needed to describe the traffic conditions:

- **Density:** It is one of the most important criteria of traffic science. It is expressed as the number of vehicles per kilometer of the road, and the maximum number of cars per kilometer is considered to be about 100 cars per lane.
- **Flow rate:** It is defined as the number of vehicles that pass through a certain part of the road per unit of time.
- **Average speed:** We define the average speed, which is given by " $u(x, t)$ ", as the ratio of flow rate to density. Average speed is also a function of time and place.

$$u(x, t) = \frac{q(x, t)}{k(x, t)} \quad (4.3)$$

Where $k(x, t)$ is for density and $q(x, t)$ is for flow rate at time " t " and at point " x ".

The traffic situation at any time is in a special state whose characteristics can be expressed using the flow rate, density, and average speed. However, there are certain traffic situations that require special attention, including the following:

- **Free-flow traffic:** When there are no other vehicles obstructing the car, the car moves at maximum speed u_f which is also referred to as free speed. Of course, this speed depends on many factors, including road design, obstacles, weather, etc. At

free speed, density and flow rate are close to zero.

- **Saturated traffic:** In saturated streets, the flow rate and speed are close to zero, vehicles are in a queue, and the density has its maximum value, k_j .
- **Maximum capacity traffic:** The capacity of a road is equal to the value of the maximum flow rate (q_c). For the maximum flow rate value, there are also corresponding values of speed at maximum capacity (u_c) and density at maximum capacity (k_c). It should be noted that the speed at maximum capacity (u_c) is lower than the free speed (u_f).

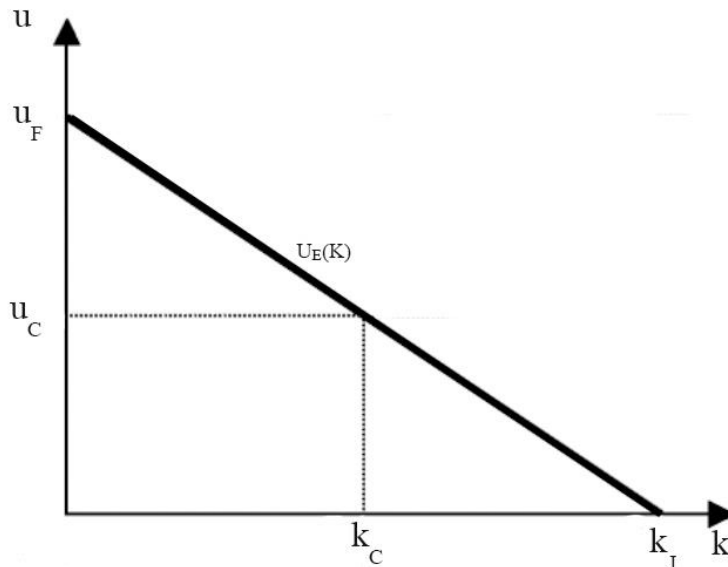


Figure 4.2. The fundamental diagram of traffic theory according to GreenShield.

The variables of boundary conditions are of particular importance, especially the speed at maximum capacity (u_c) and density at maximum capacity (k_c). However, due to the lack of a specific formula for calculating the value of these variables and the existence of many influencing factors on their value, their calculation requires field studies. Therefore, we use a simplification to estimate the approximate value of these variables. According to the Green Shield theory, as depicted in Figure 4.2, the relationship between speed and density is a linear relationship [193], where u_c is half of u_f and k_c is half of k_j . The slope of the graph in Figure 4.2 can be calculated as follows:

$$m = \frac{u_c - u_f}{k_c} \quad (4.4)$$

After calculating the slope of the line, we can get the coordinates of every point on this line, and in this way, the speed of the vehicles can be calculated according to the density of the road segment as follows:

$$u_{estimated} = m \times (k_{current}) + u_f \quad (4.5)$$

4.2.2. Content Retrieval Mechanism

The response to the users' request can be done in four ways, which are explained below in order from the fastest method to the state with the longest delay:

- **Local response:** The users' request is satisfied by the cached content in the connected RSU. In this case, the RSU will directly respond to the request and the user will experience the least possible delay.
- **Neighbor response:** RSUs can cooperate with each other to respond to a request. If the requested content does not exist in the cache of the connected RSU, but it exists in one of the neighboring RSUs, the requested content can be downloaded from the neighboring RSU and provided to the user. Due to the local nature of the communication between RSUs, this response has a little delay and provides a user experience close to the local response.
- **Fog server response:** If the requested content is not cached in any of the RSUs, the request is sent to the Fog server, and if this server has cached the requested content, it can provide it to the RSU where the user is connected.
- **Cloud response:** If the requested content cannot be served from the cache, this request is sent to the cloud server. In this case, the user will experience a large delay.

In general, the QoE of the user mainly depends on the delay caused by the file download process. Therefore, the user's QoE can vary depending on the source of the content, where $QoE_{local} > QoE_{neighbor} > QoE_{fog} \gg QoE_{cloud}$. Local response being the fastest and having the best QoE, followed by neighbor response, fog server response, and finally cloud response.

SDN allows administrators to abstract the underlying network infrastructure for network applications and services, making it easier to direct requests to the nearest cached content. The use of SDN-based cache content dissemination in the IoV serves to minimize content retrieval time. This approach streamlines the implementation of dynamic routing policies and ensures requests are directed to the source with the least delay. Any modifications in the cached content within any RSU or the Fog server can be promptly reflected in the routing table, guaranteeing that the request is directed to the accurate source.

4.3 Proposed Caching Scheme

IoV applications are diverse and rich. However, they can be divided into two general groups: safety applications and infotainment applications. In this section, we outline a method for caching each type of content, taking into account the unique characteristics of each type.

There are four main reasons for categorizing data into safety and infotainment groups:

- **Spatial Characteristics:** The spatial usage of safety and infotainment content fundamentally differ. Safety content is typically relevant only to vehicles within the immediate vicinity of an incident; thus, it is handled locally by the RSU. In contrast, infotainment content can be demanded by vehicles across various RSUs, requiring

broader consideration.

- **Temporal Characteristics:** Safety and infotainment content exhibit different temporal characteristics. While the relevance of safety content is typically short-lived, becoming obsolete after a certain period, the popularity of infotainment content can fluctuate over time, reflecting changing user interests and trends.
- **Request Pattern:** The request patterns for safety and infotainment content are also substantially different. Infotainment content is generally transmitted to the user upon a specific request, while safety content is proactively shared by the RSU with all vehicles entering its coverage area due to its critical nature.
- **Caching Criteria:** Owing to their distinct characteristics, we apply different criteria for caching these two types of content. For instance, while content volume is an important factor for infotainment content, it can be overlooked for safety content due to its smaller volume and standardized structure. Popularity is a crucial metric for infotainment content, whereas the severity level of the incident related to safety content significantly influences the caching decision.

4.3.1. Emergency Content

Emergency contents have a high priority due to their nature, which is in line with the safety of vehicles. Caching this type of content leads to quick reception of emergency data which allows drivers to have timely and accurate responses such as braking, increasing their speed, changing direction, and so on. Of course, it should be noted that this type of content is less common than infotainment content and generally smaller in size, so caching them does not have a negative serious impact on content diversity.

Safety content, due to its unique characteristics, requires special treatment in our methodology. It possesses strong spatial characteristics, given its highly localized nature. This content is specifically relevant in the local area or region where the corresponding incident or event has taken place. For example, if an accident has occurred, the information about this accident can be useful only around the accident site, and more specifically, only on the street where the accident occurred or in the streets leading to the intersection where the safety incident occurred. We refer to the safety accident site as the **Red Zone**. Therefore, the handling of such data, in terms of caching and dissemination, is assigned to the RSU that resides within the vicinity of the incident. In this sense, an RSU located in the Red Zone independently makes the decision to cache this type of content under the STAECC framework. This approach allows us to efficiently manage and deliver safety content where it is most relevant, thus exploiting its spatial characteristics.

Safety content also displays significant temporal characteristics. After an accident, for instance, the information related to the event is particularly useful for a certain period of time. However, the relevance of this information diminishes over time. For example, after the vehicles involved in the accident have been transferred and the accident scene cleared, it becomes less useful to share the accident information. This underlines the importance of considering the temporal aspect of safety content, as it allows for efficient management of storage resources by caching the content when it's most useful and freeing up space when the content is deemed obsolete.

Based on the nature of this type of content, STAECC uses the following criteria to decide whether to cache emergency event data in the RSU:

- **Freshness:** Receiving safety events in a timely manner is critical as it can prevent

secondary incidents, such as chain accidents. However, the importance of this information fades over time and eventually becomes obsolete. It is important to note that the RSU may not have immediate access to the safety incident information. For instance, the vehicles involved in the incident may be in the blind spot of the RSU or so damaged that they cannot transmit an emergency status. However, safety messages can be piggybacked by another vehicle and delivered with a delay. To determine the freshness of the content, the RSU uses the following formula:

$$F_i = T_i - Age_i \quad (4.6)$$

Where T_i denotes the lifetime of content " i ", which is the time interval during which the content can be useful to other vehicles. It should be noted that T_i can be different for different safety events. Age_i is the time interval between the safety event occurrence time and the current time. If $Age_i < 0$ the emergency content is considered as expired and invalid.

- **Severity level:** Not all safety events have the same impact and importance. We have classified safety events into several categories " $SL = \{1, \dots, n_{SL}\}$ " which indicate the severity degree of that event. Emergency data with a high degree of severity is required to be quickly provided to other vehicles and should therefore be given higher priority for caching. The occurrence of safety events following Poisson distribution with occurrence rate for different severity levels is depicted by $\{\lambda_1, \dots, \lambda_{|SL|}\}$ where $\lambda_1 > \lambda_2 > \dots > \lambda_{|SL|}$ which means that the higher the severity level, the rarer the related safety events. The occurrence time interval between two consecutive events of severity level " i " is depicted by ΔT_n which follows an exponential distribution, $\Delta T_n \sim \exp(1/\lambda_n)$.
- **Validity level:** While timely sharing of emergency information can help improve safety and reduce risks. Sharing incorrect information can lead to the emergence of serious risks for vehicles. To reduce such risks, RSU receives information related to a safety incident from different sources (vehicles), and the more confirmation it has for a safety incident, the higher the degree of validity of the relevant information.

4.3.2. Infotainment Content

The main requests of users on the IoV are for infotainment content. Passengers can surf the web, check their social networks, play online games, or even spend time in the metaverse [194]. Infotainment content can include audio, image, and video. Quick access to this type of content can have a significant impact on the quality of users' experience. In addition, this type of content generally has a high volume and the increasing demand for infotainment content puts a lot of pressure on the vehicular network. Caching infotainment content is a reliable solution to meet these challenges. However, the storage resources in RSUs are limited, so it is necessary to provide a solution for the proper placement of content on the appropriate RSU. Also, RSUs have the possibility to cooperate with each other to respond to infotainment requests.

In this article, we propose FM3C as a framework for managing infotainment content caching which uses the following criteria to evaluate and decide on the value of infotainment content for caching.

- **Popularity:** Popularity is one of the most important criteria in content caching. If popular content is cached, it can be assured that a large number of requests can be answered locally, and conversely, caching content that is not in high demand wastes storage space and prevents more useful content from being cached. Popularity is dynamic in nature and may change over time, and as a result, calculating the popularity of content is not always a straightforward procedure.}
- **Content size:** The multimedia data used in IoV generally have a large volume. Caching a large content, if it results in responding to a large number of requests locally, leads to a reduction in the use of bandwidth and delays in receiving content. Although, on the other hand, it leads to the occupation of a large storage capacity and limits the number of contents that can be cached, reducing the variety of cached content. This is a negative criterion, which means that lower values of the content size are more desirable, and if the two contents are the same in other criteria, the content with a smaller size has a higher priority to be cached.}
- **Number of replicas:** Content can be cached on several RSUs, which leads to redundancy. Redundancy wastes storage space and reduces the diversity of cached content. On the other hand, due to the possibility of cooperation between RSUs to retrieve content with a low delay, it is necessary to avoid caching contents of which a copy is cached in another RSU as much as possible, except in special circumstances, for example, when content is extremely popular.}

For infotainment content, we have observed that the popularity of content often possesses spatio-temporal characteristics. For spatial consideration, each RSU uses a local dataset of vehicle requests within its coverage area. Therefore, we have introduced a vehicle selection mechanism for the Federated Learning process that incorporates "stay time", defined as the vehicle's residence time within the RSU's range. This is a crucial criterion in our methodology. By giving more weight to vehicles with a longer stay time in the model aggregation process, the calculated popularity of content becomes more influenced by the preferences of these local users, hence leveraging the spatial aspect.

The temporal aspect of infotainment content popularity is also considered in our methodology. Popularity trends of content can change over time. To capture these dynamics, we employ LSTM models, a type of Recurrent Neural Network. The LSTM model uses a window of past request histories to estimate the popularity of content for a horizon of future time steps, thus incorporating the temporal characteristics into our methodology.

4.4 Popularity Prediction, Based on Federated Learning

In the previous section, we introduced three criteria that are used by FM3C for the evaluation of infotainment content. Measuring the content size and number of replicas is simple and straightforward, but the same cannot be said for popularity. In many researches, popularity is calculated by counting the number of requests received for each content, but due to the dynamics of IoV, waiting for requests to be received and counted before considering content as popular is a futile approach. Therefore, we need to predict popularity and proactively cache popular content.

Accurately predicting content popularity is critical, to efficient caching in IoV. In our

proposed approach, we use a machine learning model to predict the number of content requests on the time series.

We feed the time-stamped historical requests plus features unique to the content such as movie genres as the input of the machine learning model to predict the sequence of the next popular contents. However, many users are reluctant to share such information. Additionally, sending such information to a server and training the model centrally can lead to latency, heavy processing load, and scalability issues. To overcome these challenges, we use federated learning (FL), which allows vehicles to jointly train a model using their local datasets [195]. The general optimization problem in federated learning is as follows:

$$\min F(w) = \frac{1}{k} \sum_{k=1}^{|K|} F_k(w); \quad (4.7)$$

$$F_k(w) = \frac{1}{DS_k} \sum_{i \in DS_k} \delta_i(w) \quad (4.8)$$

Where δ_i is the prediction loss for some input-output pair in the local DS dataset with model parameters " w " and " k " is the participating node index. Minimizing the value of δ_i for " k " selected nodes is equal to optimizing the value of loss for the FL protocol. In general, there are two main entities in the FL system, namely the data owners (participants) which are the vehicles in our system, and the model owner which are the RSUs. Each data owner uses its own dataset to train a local model and sends only local model parameters to the FL server. All local models collected by the model owner are then aggregated to produce a global model. The federated learning protocol consists of four phases in each training round: participant selection, global model download, local training, and model aggregation. Figure 4.3. illustrates the different phases in federated learning for IoV.

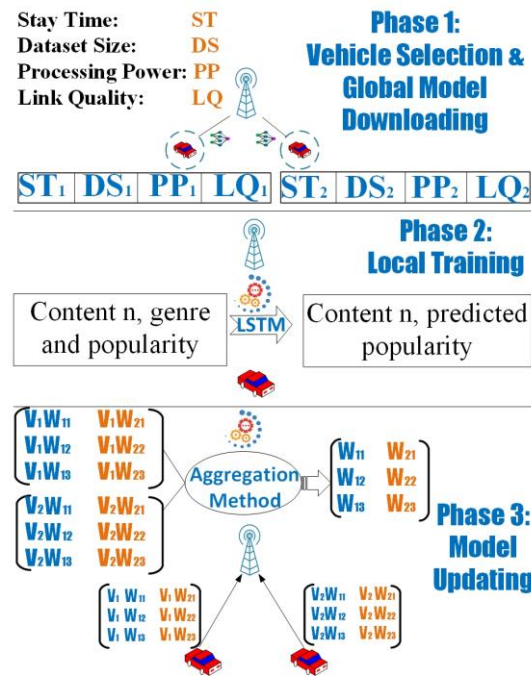


Figure 4.3. Stages of federated learning for popularity prediction in the proposed model.

4.4.1. Participant Selection

The initial stage in federated learning involves selecting the participants. In many previous works, " k " number of existing nodes were randomly selected to participate in the federated learning process without taking into account the conditions and capabilities of these nodes. In this work, we propose a method called Federated learning Participant Selection for IoV (FPSI) in which RSUs select the most suitable vehicles to engage in the learning process. In order to guarantee the model's quality, the participants must meet some expectations. As a result, RSUs implement FPSI as the selection process and consider the following criteria to evaluate vehicles.

- **Stay time:** The vehicle's residence time in the RSU range is one of the most important criteria considered for the selection of vehicles. The selected vehicles must have enough time to download the model, train it on local data, and upload the model parameters back to the RSU to ensure model quality. If the selected vehicle leaves the RSU range before this process is completed, it can negatively affect the quality of the final model. The stay time is a function of the vehicle's current location, vehicle trajectory, vehicle speed, and RSU coverage area. For RSU r_i , let v_{ij} be one of the vehicles in the coverage area of r_i , and let $\{s_m, \dots, s_n\}$, be the subset of the trajectory of v_{ij} that falls within the range of r_i . For each s_k of v_{ij} trajectory, the speed of the vehicle can be estimated using the formulas presented in the mobility model section. Then, using the following formula, it is possible to calculate the time taken by v_{ij} for traversing s_k , as given by 4.9:

$$tr_{kj} = \frac{l_k}{(u_k + N_j)} \quad (4.9)$$

Where l_k is the length of s_k , u_k is the estimated driving speed at s_k and N_j follows a gaussian distribution to demonstrate a speed variation due to the different driving behaviors. Therefore, we can calculate the stay time by adding the travel time calculated for each road segment in the route of the vehicles that are within the RSU range.

However, it should be noted that due to the dynamics of vehicle density and the dependence of the arrival time of a vehicle at a road segment on the departure time from the previous road segment, the calculated travel time error tends to be accumulated. This means that the certainty level of the estimated travel time for the first road along the route is the maximum, and for the subsequent trajectory subsets, the certainty level of the estimated travel time decreases according to the total length of the previous trajectory subsets.

Dataset size: Machine learning algorithms are fueled by data, and high-quality models can only be achieved with a sufficient amount of data. The size of the dataset should be such that it enables the discovery of relationships between the data. Consequently, a larger dataset from the selected vehicles is preferable.

- **Processing power:** Vehicles must have the necessary processing power to train the model in the allotted time. The RSU needs to receive updates from all participants before aggregating the models. As such, FL training progress is limited by the training time of the slowest participant. This is called the straggler effect [196]. The high processing power also allows the vehicle to train the model for more epochs in

a given time. As a result, the obtained model will be more accurate for predicting popularity.

- **Link quality:** The high quality of the communication link can affect the time of downloading the model from RSU as well as returning the results. According to the urban scenario considered in this article, it should be noted that due to the presence of buildings and other obstacles, the conditions of the communication link can be affected. Therefore, in the process of vehicle selection, having a stable communication link with RSU should be taken into account.

4.4.2. Global Model Download

In the subsequent phase, the chosen vehicles retrieve the base model from the linked RSU. RSUs are also responsible for setting hyperparameters, which are used to control and regulate the learning process. These hyperparameters, set by the RSU, include factors like the learning rate, the number of hidden layers, and the quantity of neurons in each layer. Another important hyperparameter to set is the weight initializer. This is a method that sets the initial values of the weights, laying the groundwork for the optimization of the neural network model. If not executed correctly, improper weight initialization can introduce issues in the learning process, or in severe cases, render learning by the model unfeasible. To perform proper initialization, we use the Xavier method [197], where the initial weights follow a normal distribution with a mean zero, and the standard deviation is calculated as follows:

$$\sigma = \sqrt{\frac{2}{n_{out} + n_{in}}} \quad (4.10)$$

Where n_{in} is the number of input neurons in the weight tensor and n_{out} is the number of output units in the weight tensor.

4.4.3. Local training

After downloading the global model from the RSU, participants start training the model using their local dataset. For this purpose, we use LSTM, which is a form of recurrent neural networks (RNNs) capable of learning “long-term dependencies” by memorizing information for long periods. LSTM deals with the vanishing and exploding gradient problem by introducing new gates, such as input, output, and forget gates. By selectively discarding or retaining information, these gates allow better control over the gradient flow and better preservation of “long-range dependencies” [198]. Before feeding the LSTM model with inputs, it is generally necessary to pre-process the input data. For this purpose, we calculate and store the statistical information of each content for each time step. Each RSU counts the number of received requests for each content such as CI in each time step and stores it as $Pop(i, t)$, also each vehicle records the history of requested contents at each time step. In addition, we do one-hot encoding for the genres of contents. In the next step, the model is fed with the historical preprocessed data for a window of previous time steps to predict content popularity for a horizon of future time steps.

The goal of the learning problem is to obtain the target weights through backpropagation of the loss function. Here we use Huber [199] as the loss which is a combination of the mean squared error function and the absolute value function. By doing so it is less sensitive to outliers in data than the squared error loss. Huber function is defined as:

$$L_{\Delta}(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{if } |y - f(x)| \leq \Delta, \\ \Delta|y - f(x)| - \frac{1}{2}\Delta^2 & \text{otherwise.} \end{cases} \quad (4.11)$$

Δ is the point where the Huber loss function changes from quadratic to linear. y is the actual popularities of the contents, and $f(x)$ is the predicted popularities of the contents.

4.4.4. Model Aggregation

After the training phase is completed, the parameters of the trained model are sent to the RSU for aggregation. To reduce communication costs and the time required for uploading the model to the RSU, the model can be compressed using methods such as subsampling, quantization, and sparsification, which approximate or encode the updates and convert them to a compressed form before sending them to the server. Once all updates are received, the RSU integrates the parameters to build a unified model, which is used to estimate content popularity within the RSU's coverage area. In subsequent rounds of federated learning, selected vehicles download and train this updated model. Most works use FedAvg, a well-known algorithm in FL, to generate the new global model w_{t+1} by averaging of all received local models w_{t+1}^k .

The standard FedAvg, used for merging models in federated learning, assigns equal significance to all received models. However, the models received may differ in quality, and for an optimal global model, models of higher quality should carry greater weight. To address this issue, we introduce the Smart Vehicular-Specific Aggregation (SVSA). To assess the quality of models received from vehicles, SVSA takes into account several criteria, such as the number of training epochs undergone by the model, the value of the loss function, the size of the vehicle's dataset, and the duration of the vehicle's stay within the RSU coverage. The first three criteria directly pertain to the quality of the trained model. The duration of stay was also considered by FPSI during participant selection, but its usage in model aggregation has a different rationale. In participant selection, the consideration of stay time was aimed at ensuring the vehicle's capacity to complete model training before exiting the RSU coverage area. However, participants' stay times can vary. The use of stay time in model aggregation is based on the premise that if more weight is assigned to vehicles with longer stay times, the computed popularity will be more reflective of these users' interests. Consequently, we can ensure that the cached content aligns more closely with the demands of the vehicles within the RSU coverage. The integration method for the received models is detailed in the following section.

$$y_k = mw_1stay_k + mw_2epochs_k + mw_3loss_k + mw_4dset_k \quad \text{where } \sum_{i=1}^4 mw_i = 1 \quad (4.12)$$

$$w_{t+1} = w_t - \eta \sum_{k=1}^K y_k w_{t+1}^k \quad (4.13)$$

$stay_k$ is the stay time of vehicle "k", $epochs_k$ is the number of epochs vehicle "k" trains

the model, $loss_k$ is the model loss value (calculated by Huber loss function) and $dset_k$ is calculated as $d_k/|d|$ in which $|d|$ is the sum of the dataset sizes for all vehicles involved in federated learning, and d_k is the dataset size of the $k - th$ vehicle. η is the learning rate, w_t is the matrix of weights in the previous round and w_{t+1}^k is the matrix of weights received from the $k - th$ vehicle.

Finally, after obtaining the aggregated model, RSU can estimate the future popularity of contents in its coverage area.

4.5 Multi Criteria Decision Making

In the previous sections, the method of evaluating emergency and infotainment contents for caching was examined. We have considered different criteria for each type of content, which allows us to check the usefulness of caching a content from different aspects. Also, in the integration of models in federated learning, we introduced four criteria. The existence of several criteria with different degrees of importance and scales leads to a multi-criteria decision-making (MCDM) problem. MCDM provides robust decision-making in domains where choosing the best alternative is very complex. Many decisions are made based on different criteria, so a decision can be made by determining the weight of different criteria. For this purpose, we use the AHP, which is a form of multi-criteria decision-making mechanism that helps the decision-making process by organizing and analyzing complex problems in the presence of conflicting criteria. AHP calculates a weight for each criterion using their relative importance. For this, we consider a matrix as follows:

$$\begin{matrix} & m_1 & m_2 & \cdots & m_n \\ m_1 & r_{11} & r_{12} & \cdots & r_{1n} \\ m_2 & r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_n & r_{n1} & r_{n2} & \cdots & r_{nn} \end{matrix} \quad (4.14)$$

In the above matrix if $i = j$, $r_{ij} = 1$ otherwise $r_{ij} = 1/r_{ji}$. m_i represents different criteria and the relative importance of each criterion m_i compared to m_j is represented by r_{ij} . In addition, it is worth noting that these relative values can be adjusted by experts to match the requirements of different scenarios without requiring further changes in the proposed algorithm. After determining the relative importance of the criteria, the weight of each metric W_m is calculated as follows:

$$W_m = \frac{1}{n} \sum_{i,j=1}^n \frac{r_{ij}}{\sum_{i=1}^n r_{ij}} \quad (4.15)$$

After calculating the weights, before using them, for example, to calculate the usefulness of caching a content, it should be noted that the criteria have different units and scales. For example, stay time is in seconds while processing power is in megahertz. Subsequently, before combining different criteria, normalization should be done. For this purpose, we use the 1-norm method. Then, by obtaining the sum of the product of each criterion in its weight, we can reach a single value and make our decisions based on this value.

4.6 Global Content Placement

Considering the local nature of the emergency contents, a decision is made to cache the safety content in the RSU located in the Red Zone (safety event site). However, for infotainment content, caching decisions are made centrally at the Fog server to ensure efficient collaboration among RSUs in response to requests

After calculating and ranking the contents based on popularity in each RSU, it is the task of the Fog server to globally select the appropriate contents to be cached in each RSU. The criteria explained in the section related to infotainment content are the basis for the decision by the Fog server on this matter. Based on the mentioned criteria and the weight of each criterion determined by the multi-criteria decision algorithm, the Fog server calculates the desirability of caching different contents for each RSU and sorts them in descending order. Each RSU first adds the most desirable content to its cache list. If a content has the same desirability rank in multiple RSUs at the same time, the RSU with the highest desirability score for the content will add the content to its cache list. One of the goals of global decision-making for caching is to reduce redundancy, and this is achieved by considering the number of copies of each content in the decisions. As a result, after a content is added to the cache list in one of the RSUs, the desirability of caching that content for other RSUs is recalculated (decreased).

With the information it receives from all RSUs and its global view of the network, Fog server can determine the optimal contents to cache in an efficient way for all RSUs, avoiding storage space waste and redundancy. The process of globally selecting contents to cache in RSUs can be repeated at specified time intervals or whenever the cache hit rate drops below a specified threshold, at which point new contents are selected to cache. Also, due to having a storage space, the Fog server caches contents with a high degree of usefulness that are not cached on any RSU. The Fog server is the last stop of requests before sending them to the Cloud, and if the requested content is cached in this server, the request can be answered with a relatively lower delay than the Cloud.

4.7 Performance Evaluation

The performance of the proposed algorithms was evaluated through simulations of an urban IoV environment consisting of 5 RSUs and a Fog server. Considering the different natures of infotainment and emergency contents and the difference in the proposed methods for caching each, we have implemented two different simulations for evaluating the performance of the proposed algorithms.

4.7.1. Infotainment Content Caching

For infotainment contents, we have used the real-world MovieLens dataset [200] for our experiments. MovieLens 100k contains 100,836 ratings from 610 users on 9,742 movies. Each user rated at least 20 movies. Users' contextual information, e.g., gender, age, address and occupation, is also provided in the dataset. To simulate the content requests, we assume that movie ratings are the same as user requests for content. In our simulation, MovieLens users correspond to vehicles, movies with contents, and the recording time of the movie rating to

the content request time. The paths traveled by vehicles are created using the SUMO simulator and are used in the simulation after processing. We implement the caching strategy by the Python programming language, and TensorFlow is used for the machine learning parts.

To evaluate the efficiency of the proposed method, we compared FMC with the following baseline caching schemes:

- **Random:** Contents are randomly selected by the Fog server for caching. To improve its performance, we assumed the avoidance of caching duplicate contents in neighboring RSUs.
- **LRU:** When the storage capacity limit is reached, it first deletes the least recently used content in the cache to make room for new content.
- **Intelligent Caching Strategy [157]:** This method leverages the spatio-temporal characteristics of data, categorizing it into various types. Particularly for infotainment content, the strategy takes into account the popularity level of the content. Unlike our method, which focuses on caching content at the nodes (RSUs), the Intelligent Caching Strategy opts to cache contents directly on the vehicles. For the purpose of a fair comparison with our method, we assume that the total storage capacity of the vehicles in this strategy is equivalent to the cumulative storage size of the RSUs in our approach.
- **Centralized:** In this method, RSU trains a centralized model to predict content popularity. Most users are likely to have made multiple requests for different contents before entering the range of the current RSU. However, in the centralized approach, the RSU has no access to the users' request history and only relies on anonymously recorded history of previous requests.
- **Oracle:** It has detailed information about the future requests of users. Oracle provides the best performance and cannot be beaten by any other method. The purpose of this method is to provide an upper bound for the performance of other algorithms.

The most important criterion for checking the performance of a caching method is the Cache Hit Rate (CHR), which is defined as follows:

$$CHR = \frac{CH_{local} + CH_{neighbor} + CH_{fog}}{request_count} \quad (4.16)$$

In the expression, CH_{local} denotes the number of requests satisfied by associated RSUs, $CH_{neighbor}$ denotes the number of requests fulfilled by neighbor RSUs and CH_{fog} is the number of requests that use cached contents in the Fog. $request_count$ is the total number of requests made by users.

CHR is the main criterion in the sense that other measures such as the delay, QoE perceived by users, the utilization rate of backhaul communication, etc. are closely related to the cache hit rate and they can be calculated according to the cache hit rate. In Figure 4.4, we examine the performance of different methods when using different storage spaces:

The results depicted in Figure 4.4. demonstrate that the proposed FM3C method closely mirrors the Oracle's performance and surpasses other reference caching schemes across all storage sizes in terms of CHR. The difference between FM3C and other baseline performances is particularly pronounced at lower storage sizes. This is because as the cache size expands, the probability of a cache hit increases due to a larger amount of content being actively cached in the RSUs. Additionally, it's worth noting that the central method's performance is the closest to FM3C and the random algorithm's performance improves linearly with the increase

of storage space. Conversely, due to the dynamism of the vehicular environment, the LRU algorithm registers the poorest performance. In smaller storage sizes, the performance of the intelligent caching strategy is superior to methods such as random and LRU, which do not consider data characteristics like popularity. However, in larger storage sizes, due to the lack of a clear mechanism for communicating and sharing content between different areas using infrastructure, its performance experiences the least improvement with the increase of storage sizes.

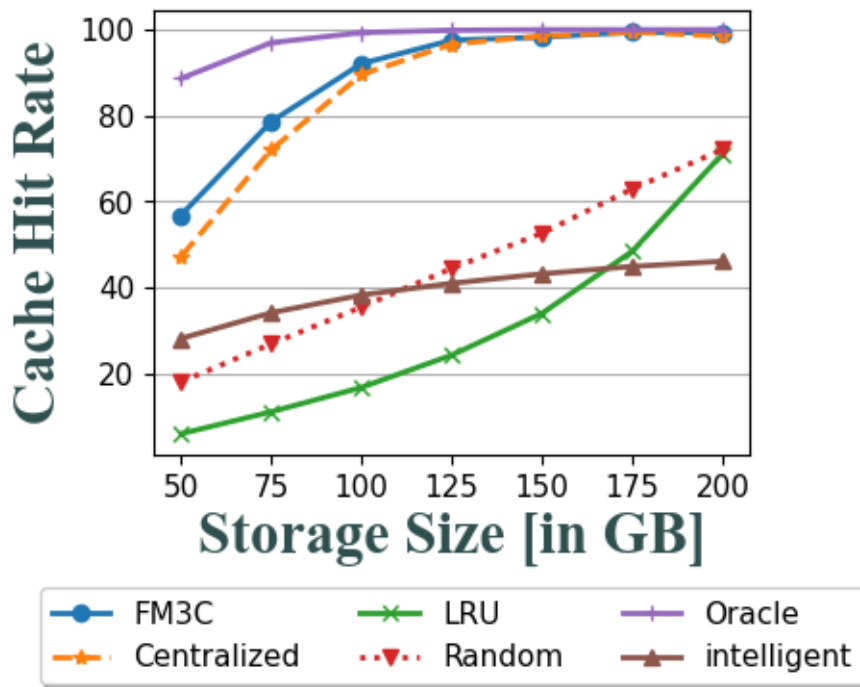


Figure 4.4. Performance comparison between FM3C and: (i) Oracle, (ii) Centralized, (iii) Intelligent Caching Strategy, (iv) LRU, and (v) Random, in

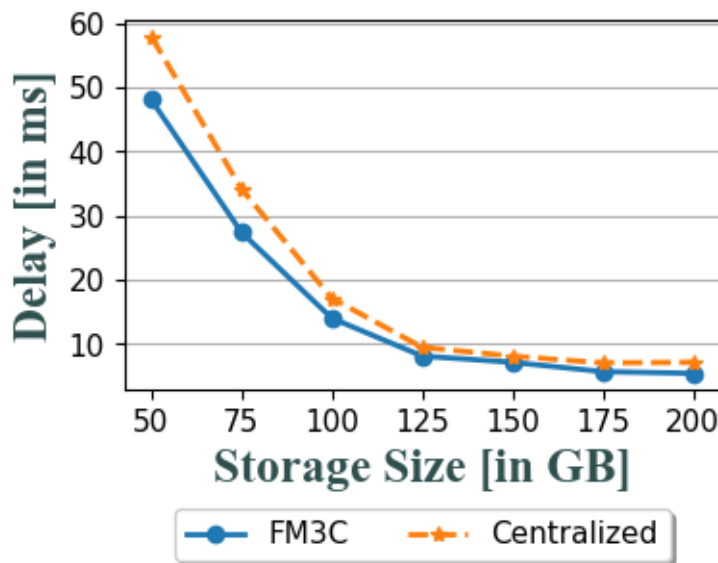


Figure 4.5. Comparative analysis of response delay between FM3C and the centralized method across varying storage sizes.

In our system, there are four methods to respond to requests, three of which are related to retrieving information from the cache. Retrieving information from the cache in all three methods is much faster than retrieving information from the cloud, but the delay in accessing cached content varies depending on where it is stored. For example, if the requested content is stored in the associated RSU, we will experience the fastest response possible. However, if we need to retrieve information from neighboring RSUs, we will experience a relatively higher delay, and retrieving content from the Fog results in even longer delays. Since content access time is considered the most important factor in measuring users' QoE, in addition to cache hit rate, the cache hit location should also be considered. Therefore, in Figure 4.5. and Figure 4.6, we compare the delay and breakdown of cache hit rate of FM3C with its closest contender, the Centralized method.

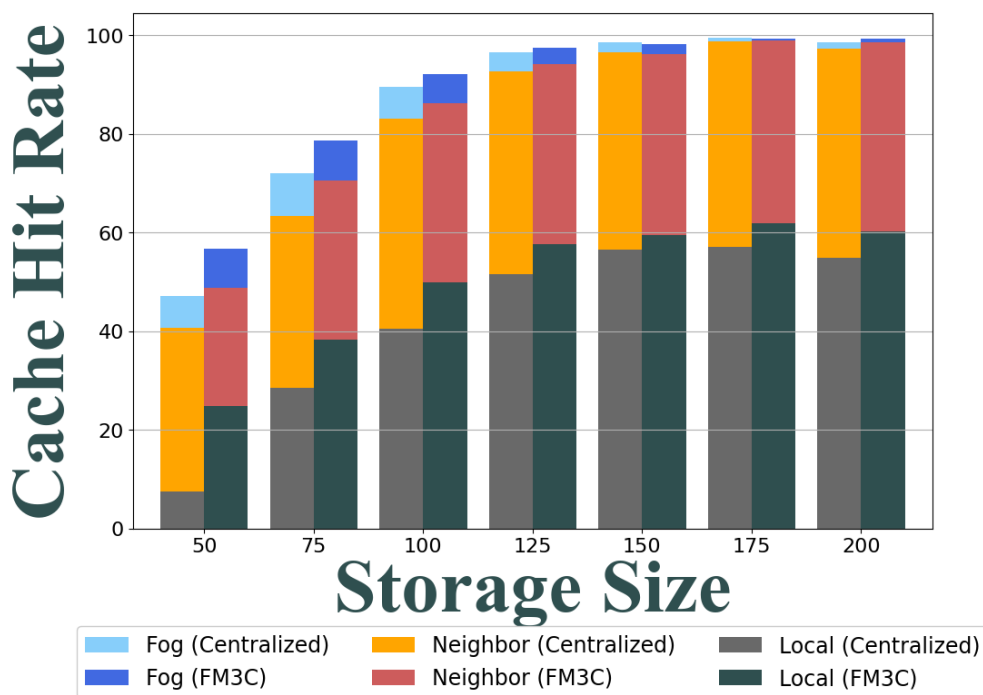


Figure 4.6. Disaggregated cache hit rate comparison between FM3C and Centralized method, in terms of different storage sizes at the RSUs.

As depicted in the diagram, the majority of cache hits for FM3C occur locally. Conversely, with the Centralized method, particularly in the case of limited storage capacity, most cache hits stem from content cached in adjacent RSUs. This outcome is anticipated as the FM3C method trains its model in a federated manner, utilizing data from vehicles within its coverage area. This process ensures that the model aligns more closely with the current user interests, enhancing the likelihood of caching content that will be requested by users in its coverage area in the near future. Meanwhile, the Centralized method trains its model solely based on the request history stored in the RSU, which might not accurately reflect the interests of the current users within its range. Therefore, while the Centralized method exhibits comparable performance to FM3C in terms of cache hit rate, FM3C enhances QoE by storing content as proximate to the users as possible. This outcome is also evident in Figure 4.5, where FM3C incurs less delay across all storage sizes compared to the Centralized method, thereby enhancing the QoE for users utilizing FM3C.

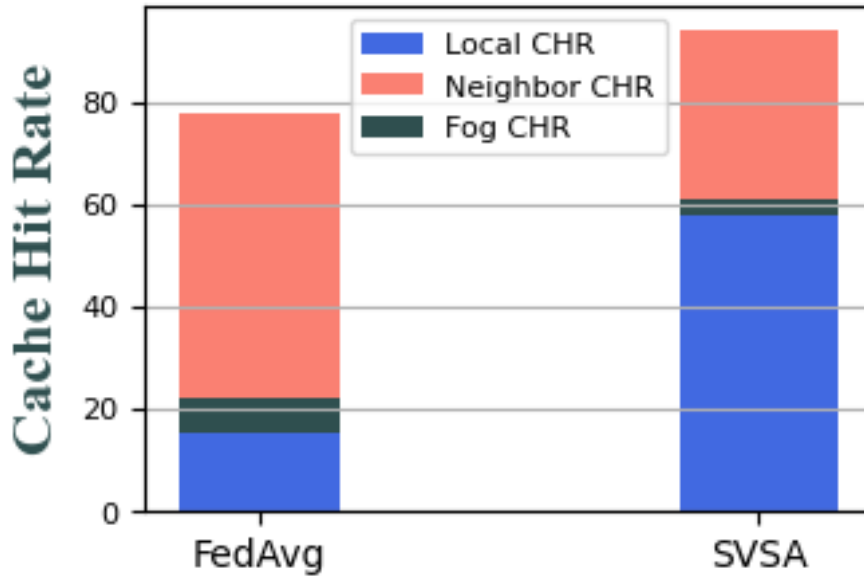


Figure 4.7. A comparison of aggregation methods in Federated learning between SVSA and FedAvg.

An important aspect of our work is the development of a novel mechanism for aggregating federated learning models called SVSA, which assigns weights to each model based on their accuracy and quality. To evaluate the effectiveness of SVSA, we compared the performance of the FM3C model when using SVSA versus FedAvg, a commonly used aggregation method in federated learning. To do so, we implemented FM3C twice, once with SVSA and once with FedAvg, while keeping all other steps in the FM3C constant. By comparing the performance of the two versions of FM3C, we were able to directly assess the impact of the aggregation method on the model's accuracy and quality. As demonstrated in Figure 4.7, the proposed aggregation mechanism outperforms FedAvg. To fully understand this advantage, it's important to focus on the local cache hit rate rather than the overall cache hit rate. The local cache hit rate reflects the RSU's ability to accurately predict popular and relevant content for the next time frame, with the aggregation mechanism playing a critical role in this prediction. The variety of cached content has increased at the neighbor level due to fog orchestration and efforts to avoid duplicates, leading to a more satisfactory neighbor cache hit rate even if the popularity of content may not be accurately predicted for the next time frame. In conclusion, the superiority of SVSA in the local cache hit rate highlights the effectiveness of the proposed aggregation mechanism, while the relative compensation of this advantage by FedAvg in the neighbor cache hit rate demonstrates the impact of the fog-based cache orchestration.

Another key contribution of our work is the development of a novel vehicle selection mechanism for federated learning called FPSI. To assess the effectiveness of this mechanism, we compare the performance of FM3C using FPSI with a scenario where vehicles are randomly selected for participation in federated learning. As shown in Figure 4.8, the proposed vehicle selection mechanism has a substantial impact on the efficiency of the caching mechanism. Without an appropriate participant selection mechanism, the caching method's performance is not satisfactory, resulting in a model that inaccurately estimates the content popularity and selects inappropriate contents for caching. Additionally, we found that when participants are randomly selected, the FM3C's performance is similar to that obtained by randomly selecting content for caching

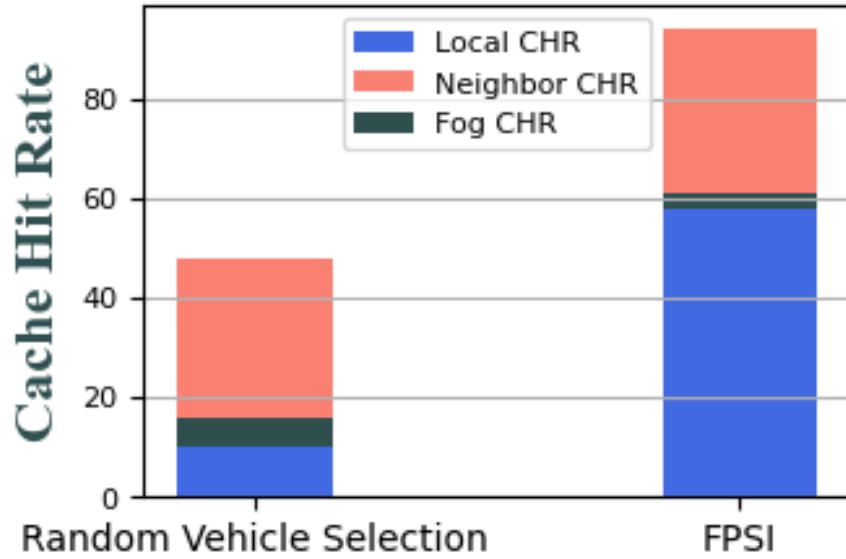


Figure 4.8. A comparison of participant selection methods in Federated learning between FPSI and random selection.

4.7.2. Emergency Content Caching

In this section, we evaluate the proposed method for caching safety-related contents. Given the spatio-temporal characteristics of this type of content and its localized impact, each RSU independently decides to cache the contents related to safety events that occurred in its area. Therefore, in this section, we consider only one RSU for this evaluation.

We have designed a dataset specifically for evaluating the proposed method for emergency content caching. The incidents related to safety in this dataset are modeled to follow a Poisson distribution, and their rates vary based on the severity level of the incident. Emergency contents are typically small in size as they are informative and must be quickly distributed. Therefore, in our evaluation, we have assumed the size of all emergency contents to be 1MB.

Before examining the performance of the STAECC, we first examine the effect of storage space on the average cacheable contents. Figure 4.9 illustrates the proportion of emergency content that can be cached over time with varying amounts of storage space. This proportion is not dependent on a specific caching algorithm and serves as a baseline for future results. Unlike infotainment contents, emergency contents are actively sent to vehicles entering the RSU's range rather than being requested, as they are crucial for taking precautionary measures. Hence, instead of calculating the cache hit rate, the availability of appropriate content and storage allocation method are considered. For the results of the next sections, we have considered a storage space of 30 MB. As shown in Figure 4.9, with this storage, 73% of the produced safety-related contents can be cached on average.

In our proposed caching method STAECC, in order to select the appropriate emergency contents for caching, three criteria of freshness, intensity and validity are taken into account. In the following, we will separately evaluate the impact of each criterion on caching.

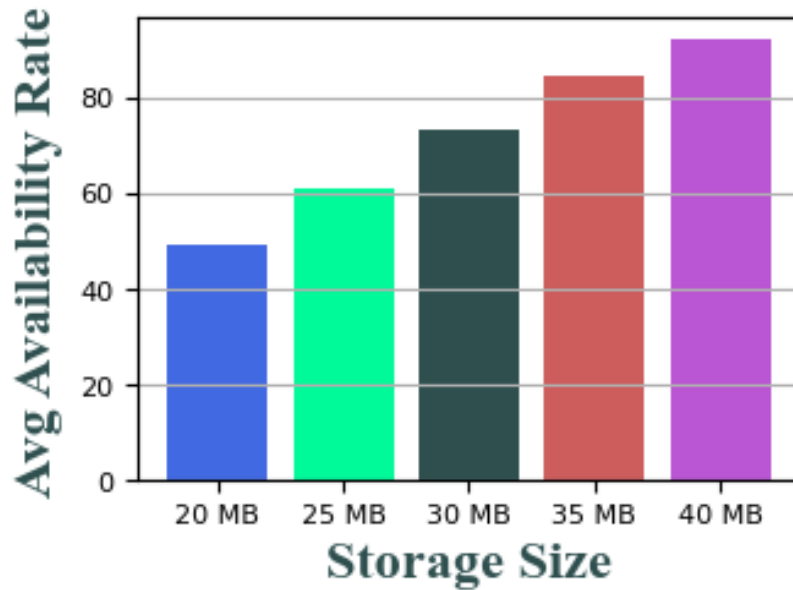


Figure 4.9. Content availability rate in terms of different storage sizes at the RSUs.

Figure 4.10. illustrates the impact of freshness on the propensity for caching content. The red portion of the bar chart represents the number of emergency contents generated in the current time step. Conversely, the blue portion signifies the number of these newly formed contents chosen for caching. It can be observed that STAECCE acknowledges freshness and aims to prioritize and cache new content, particularly when the volume of recently generated content is not substantial. However, when there is a surge in newly generated content, STAECCE becomes more selective, simultaneously considering additional criteria such as validity and severity to allocate the storage space to more valuable contents.

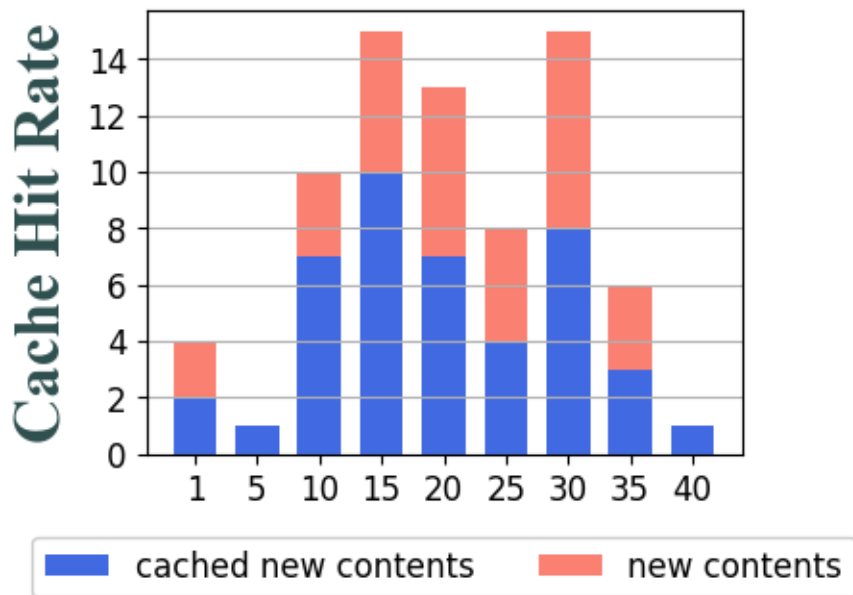


Figure 4.10. Examining the ratio of cached contents in terms of freshness.

The critical factor in selecting the appropriate emergency contents for caching in our proposed STAECCE method is the severity of the safety incident that the content is related to.

We have considered five levels of severity for emergency messages, with level five being dangerous and extremely important events, and level one being more common incidents. As depicted in Figure 4.11, the likelihood of caching the content increases with the increase in the severity level of the safety incident, and almost all contents with severity levels four and five are stored in the cache. This implies that drivers can always have access to information about crucial safety incidents.

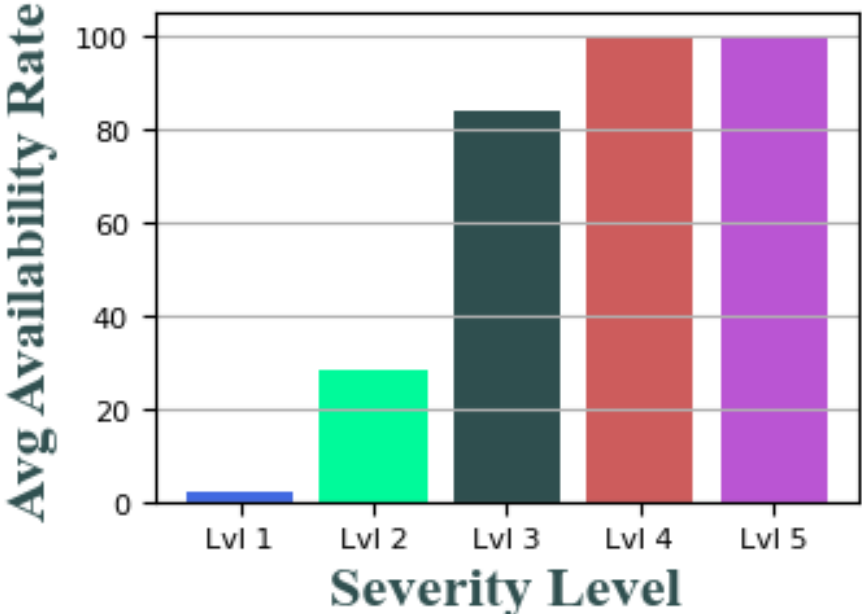


Figure 4.11. Examining the ratio of cached contents in terms of severity level.

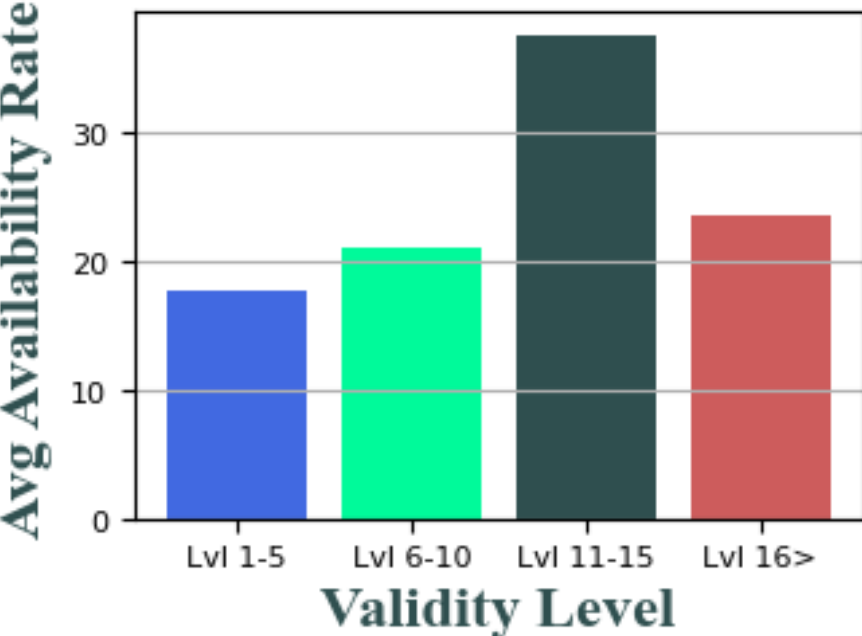


Figure 4.12. Examining the ratio of cached contents in terms of validity level.

The validity level is the final factor that we evaluate in our proposed STAECC method. As shown in Figure 4.12, the proposed caching mechanism demonstrates that the higher the

validity level of emergency messages, the higher the probability of caching the related emergency contents. It is worth noting that the number of cached contents with more than 16 confirmations is lower compared to the previous level, and this is due to the low probability of receiving such a high number of confirmations for an emergency message. In other words, in the scenario implemented, it is rare to receive these many confirmations for an emergency content.

4.8 Conclusion

In this chapter, considering the different natures of emergency and infotainment contents, we have proposed two innovative caching methods, STAECC and FM3C. The STAECC method was designed specifically for the unique characteristics of emergency content, prioritizing the allocation of storage space to the most important and valid safety information. This increases the availability of these contents for vehicles within the RSU range. On the other hand, FM3C was proposed to improve the user experience of infotainment contents through a three-step process. The first stage involves a federated learning method to predict content popularity with user privacy protection, the second stage combines the predicted popularity with other characteristics through a multi-criteria decision-making method, and the third stage involves global decision-making and orchestration by the Fog server. The simulation results show the remarkable performance of both the STAECC and FM3C methods.

In the next chapter, we aim to extend our methods to more realistic scenarios that may include areas without RSUs and to explore incentive mechanisms for vehicles to participate in the federated learning process and content caching.

5. Enhancing IoV with SONA: A Synergistic Approach to UAV Deployment and Blockchain-Based Caching

5.1 Introduction

The recent surge in the evolution of Intelligent Transportation Systems (ITSs) is significantly characterized by the advent of the Internet of Vehicles (IoV). This transition heralds a new phase of intelligent, interconnected transportation, necessitating advanced solutions to enhance network services and improve user experiences. For clarity and ease of understanding, this article will use the terms 'vehicles' and 'users' interchangeably.

The rapid expansion of IoV technologies has catalyzed the development of numerous applications, significantly enhancing vehicular networks. These advancements span a broad spectrum, from traffic information and path planning to safety mechanisms and beyond. Notably, the domain of in-vehicle entertainment is set to experience substantial growth. The advent of autonomous vehicles introduces a paradigm shift, liberating passengers from the demands of driving and opening avenues for immersive experiences through multimedia content, social media interactions, the Metaverse, and virtual reality. Projections indicate that by 2040, autonomous vehicles will account for approximately 75% of all road traffic, potentially reaching a market valuation of \$7 trillion by 2050 [201].

While the developments in the IoV landscape are promising, they present two formidable challenges that must be addressed to ensure high-quality service for IoV users. The first challenge pertains to Quality of Service (QoS) and the need for low delay. Many IoV applications, especially those focusing on safety, necessitate stringent delay requirements. Furthermore, entertainment-related applications not only demand low latency but also consume extensive bandwidth, potentially straining the backhaul network if all content were to be fetched from cloud servers. An effective solution to mitigate this challenge is content caching. By caching popular content closer to the users, it is possible to significantly reduce redundant data transfers and enhance transmission efficiency.

The second significant hurdle is ensuring adequate coverage and data rates to meet the diverse demands of IoV applications and user requests. The existing terrestrial infrastructure often falls short of fully spanning the vastness of road networks, a situation further complicated by natural disasters or when systems are overwhelmed by exceptionally high volumes of requests. Deploying Unmanned Aerial Vehicles (UAVs) emerges as a flexible and swift solution to augment existing infrastructure, making UAVs an integral component of modern wireless communication systems. Their deployment can extend coverage to remote and inaccessible areas and provide additional capacity where it is most needed, thus addressing both coverage and data rate challenges.

The efficacy of caching in the IoV framework is underpinned by a simple yet profound observation: statistical data reveals that users within the same geographical vicinity often exhibit similar content preferences and frequently request identical files [202]. By strategically caching this popular content closer to the users, it becomes possible to significantly reduce repetitive data transfers and shorten response delays. This direct fetching of requested content from a local cache, rather than a distant server, markedly enhances transmission

efficiency and user experience.

However, the implementation of an effective caching strategy is not without its challenges. Given the limited resources available for caching in terms of storage not all content can be cached. The key, therefore, lies in being selective, focusing on caching content that is truly popular and in demand. This necessitates a predictive approach to caching, where the future popularity of content (and hence the predicted number of requests) is estimated based on historical user request patterns. Yet, this predictive model faces a significant hurdle: users' reluctance to share their content request history, owing to privacy concerns. Such reluctance can impede the accuracy of content popularity predictions, a cornerstone for effective caching.

To navigate this challenge, our methodology employs Federated Learning, innovatively modified to enhance participant selection and model aggregation. This approach allows for the aggregation of insights from multiple users' data without needing to directly access or centralize this sensitive information. By doing so, we ensure a privacy-preserving mechanism for predicting content popularity, crucial for real-time, efficient data access in IoV environments. This nuanced, adaptive strategy not only respects user privacy but also optimizes the allocation of limited caching resources, ensuring that only the most relevant and demanded content is cached and readily available to users.

We further address the overarching challenges within the IoV ecosystem by turning our attention to the deployment of UAVs in vehicular networks, which introduces a suite of transformative benefits that address the key limitations of current terrestrial infrastructure.

1. **Emergency Infrastructure:** UAVs can provide temporary communication infrastructure during natural disasters, vandalism, or scheduled infrastructure repairs, ensuring uninterrupted network service.
2. **Traffic Offloading:** They play a crucial role in offloading vehicular traffic from dense cellular networks, enhancing wireless capacity and coverage, especially during temporary events or in hotspot areas.
3. **Extended Coverage:** UAVs extend coverage to remote and inaccessible areas, where conventional wireless services fall short.
4. **Edge Computing:** Serving as mobile edge computing servers, UAVs support enhanced IoT and IoV functionalities, facilitating real-time data processing and analytics closer to end-users.

Despite these advantages, the deployment of UAVs is not devoid of challenges. The cost of deploying a comprehensive multi-UAV network, capable of catering to all ground users across extensive operational areas, necessitates significant capital investment. Additionally, unlike terrestrial Roadside Units (RSUs) with access to an uninterrupted power supply, UAVs are battery-operated, limiting their operational duration. Moreover, the strategic 3-D placement of UAVs to maximize connectivity and throughput—while ensuring optimal service to the greatest number of vehicles—introduces a layer of complexity in their deployment.

Addressing these challenges, our methodology, dubbed Strategic Optimization for Networked Aerial-vehicles (SONA), embarks on optimizing both the number and the placements of UAVs. This optimization hinges on accurately gauging current and predicted vehicular density and demand, including data rate requirements. To achieve this, we leverage Long Short-Term Memory (LSTM) models for predicting vehicular traffic patterns within specific areas. Building upon these predictions, we develop an optimization model aimed at maximizing throughput and QoS with the minimal deployment of UAVs, thereby ensuring efficient energy consumption. Crucially, our approach capitalizes on air-to-ground communication channels, enhancing the probability of establishing line-of-sight links, which

significantly improves channel gain and reduces power consumption. In the dynamic and ever-evolving landscape of vehicular networks, we frame the optimization challenge as a Markov Decision Process (MDP). Through the application of Deep Reinforcement Learning (DRL) techniques, we derive an optimal deployment strategy that adeptly navigates the complexities of UAV utilization in IoV environments.

By integrating these advanced predictive and optimization models, SONA sets a new precedent for UAV deployment in enhancing IoV infrastructure, offering a systematic approach to overcoming the prevalent challenges of cost, coverage, and connectivity.

Central to the advancement of the IoV systems is the challenge of decision-making regarding content caching and UAV deployment. Traditional approaches often rely on centralized control, which poses several questions: Who determines what content to cache? How is UAV deployment decided? Is it feasible for individual nodes, such as RSUs or UAVs, to make these decisions autonomously? The effectiveness of IoV systems significantly depends on the cooperative behavior between RSUs and UAVs. This cooperation is not only pivotal for optimizing caching strategies but also for making informed decisions about UAV deployment. By working together, RSUs and UAVs can circumvent the pitfalls of redundant caching and ensure a more diverse and efficiently distributed content cache. Similarly, collaborative decision-making can streamline UAV deployment, maximizing coverage and service quality while minimizing operational costs.

Our strategy leverages blockchain technology to facilitate a decentralized, immutable, and transparent framework for RSU cooperation. This blockchain-enabled system allows RSUs to independently make informed decisions on optimal caching locations, promoting content diversity and availability without centralized oversight, perfectly aligning with the distributed nature of IoV networks. Unlike centralized decision-making systems, blockchain technology offers substantial benefits: it eliminates single-point-of-failure risks, obviates the necessity for establishing and maintaining a centralized entity, and inherently supports the decentralized architecture that is fundamental to many IoV applications.

For UAV deployment, however, the role of blockchain extends to orchestrating a consensus among all nodes involved. Instead of autonomous decisions by individual nodes, blockchain facilitates a collective consensus mechanism for UAV deployment. This ensures that deployment strategies are not determined in isolation but are the result of a collaborative decision-making process, reflecting the collective intelligence and current needs of the entire network. By recording and sharing data on content popularity, vehicular movements, and network demands on the blockchain, all participating nodes—RSUs and UAVs alike—can access real-time information to reach a consensus on where and when UAVs should be deployed. This method enhances the network's responsiveness and adaptability, significantly improving scalability and security. Through this consensus mechanism enabled by blockchain, decisions regarding UAV deployment are optimized for network resource utilization and user experience, marking a significant advancement in the management of IoV infrastructures.

Our Contributions

This research addresses critical challenges in the domain, focusing on enhancing connectivity, service quality, and paving the way for innovative applications. The contributions of this study are both foundational and transformative, offering solutions that blend advanced computational models with practical, real-world applicability. Specifically, we present:

- **Comprehensive UAV Mobility and Transmission Models:** We have developed detailed models for UAV mobility and transmission. These models serve as a cornerstone for designing energy-efficient UAV networks, ensuring extended operational times and

greater coverage.

- **Privacy-Preserved Content Popularity Prediction:** Introducing a novel mechanism for content popularity prediction, our approach respects user privacy by employing advanced federated learning techniques. This mechanism accurately forecasts content demand without direct access to individual user data, facilitating efficient and privacy-compliant content caching strategies.
- **Blockchain-based, Decentralized Caching Mechanism:** We propose an innovative caching strategy that utilizes blockchain technology for secure, decentralized decision-making. This strategy empowers RSUs and UAVs to make collaborative caching decisions, optimizing content placement and minimizing redundancy, thereby enhancing the network's efficiency and responsiveness.
- **Pareto Optimization for UAV Deployment:** Our research introduces a Pareto optimization approach to UAV deployment, balancing the trade-off between operational costs and user satisfaction. This optimization framework ensures that UAV resources are utilized in the most effective manner, delivering high-quality services while minimizing the financial and environmental footprint of UAV operations.
- **MDP-based Problem Transformation:** We have transformed the optimization challenge into an MDP, allowing for a dynamic analysis of UAV movement and content request patterns. Employing DRL, our model adapts to evolving network conditions, optimizing UAV deployment and content caching in real-time to meet the changing demands of IoV users.

The structure of this article is designed to provide a coherent progression through the topics under discussion. Section 2 begins with a review of the existing literature, setting the stage for our contributions. This is followed by Section 3, which offers a detailed explanation of our comprehensive system model. In Sections 4 and 5, we introduce our innovative caching algorithm and UAV deployment strategy, respectively. The results of our simulations are presented in the subsequent section, demonstrating the efficacy of our approaches. We conclude the article by summarizing our key findings and exploring potential avenues for future research.

5.2 System Model

Our system model introduces an innovative approach tailored for the vehicular network, comprising several key components that collaboratively facilitate seamless communication and content delivery. It is important to clarify that the order in which these components are discussed does not reflect their importance or hierarchy within the network. The primary elements are:

Vehicles: Serving as the primary users, vehicles play a crucial role in our network. Each vehicle follows a trajectory that intersects with various RSUs or UAVs. The connectivity needs and content request of a vehicle at any given point are satisfied by either an RSU or a UAV, depending on which one it is linked to. The demand for content is influenced by factors such as user preference, the nature of the journey, and advancements in vehicular technology.

Content: Our network supports a wide variety of content, with a focus on multimedia for in-vehicle entertainment. The rise of autonomous vehicles amplifies the need for diverse entertainment options during travel, reflecting an increased demand for engaging side activities.

Road Side Units (RSUs): RSUs provide connectivity to vehicles within their coverage area.

An essential feature of the RSUs is their built-in storage that facilitates content caching. By storing content, the RSUs significantly enhance the QoS by reducing content retrieval latency. RSUs also feature an ability to collaborate with other units to augment the QoS and fulfill content requests more efficiently.

Unmanned Aerial Vehicles (UAVs): UAVs offer a cost-effective solution to extend network coverage, particularly in areas without RSU support. Serving as mobile RSUs, they not only broaden the network's reach but also possess storage capabilities, albeit typically less extensive than those of RSUs, for content caching. UAVs prove invaluable in regions of high vehicle density, where a single RSU may be insufficient to provide satisfactory QoS.

The network architecture, illustrated in Figure 5.1, demonstrates the interconnectedness of these components in delivering efficient and reliable service. This model underscores the potential of integrating vehicles, RSUs, UAVs, and blockchain technology to meet the dynamic needs of vehicular network users.

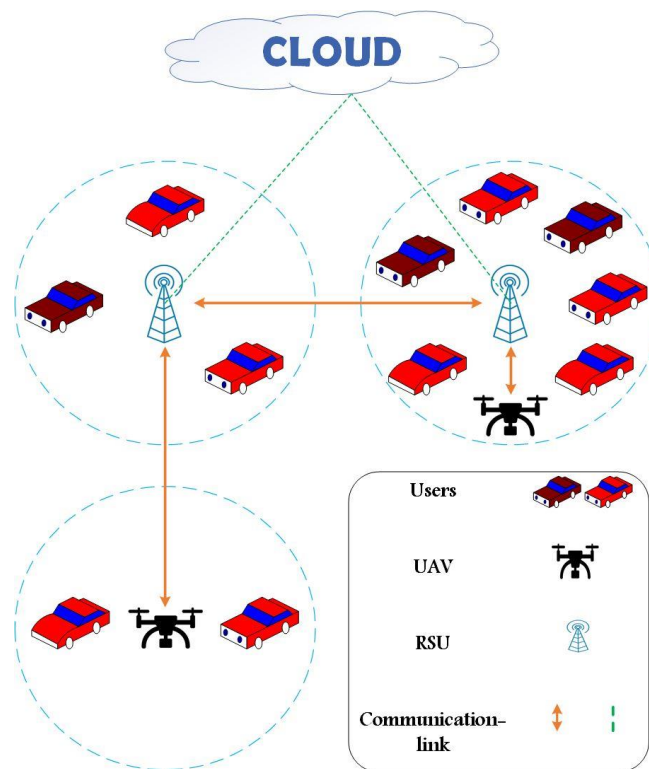


Figure 5.1. The proposed vehicular network architecture.

5.2.1. Caching Model

The caching model is pivotal in determining the QoS experienced by users in our vehicular network system. As vehicles navigate, they may request various types of content. It is crucial to underscore that at any given time slot " t ", a vehicle (user) is limited to requesting a single piece of content. Rapid fulfillment of these requests is essential for enhancing QoS, underscoring the value of an effective caching strategy.

Upon receiving a content request, the connected node (either an RSU or UAV) checks its local storage to ascertain the presence of the requested content. If available, the node immediately fulfills the request, thereby minimizing delay. Conversely, if the content is not locally cached, the node queries other nodes to identify if any have cached the requested

content. Retrieving content from another node incurs a delay, albeit shorter than sourcing the content from the cloud, which represents the longest delay scenario.

The decision to cache content at each node is governed by several criteria, with predicted content popularity being paramount, as it often dictates whether or not the content should be cached. The relevant information about each content, including its predicted popularity, is recorded on the blockchain. This ensures transparency and allows all nodes to access critical data for informed decision-making.

After these details are committed to the blockchain, each node evaluates the appropriateness of caching specific content based on caching suitability scores. This evaluation encompasses a comparative analysis of the scores across all nodes, including the evaluating node itself. Should the evaluating node possess a higher suitability score for caching a particular content, it proceeds to cache the content. Conversely, if another node demonstrates a higher suitability score, the evaluating node refrains from caching, thereby optimizing the overall caching strategy and ensuring efficient use of storage capacities across the network.

5.2.2. User Satisfaction Model

To quantify the Quality of Service (QoS) for users, we introduce a user satisfaction model in our system that centers around data rate, a key factor crucial for enhancing user satisfaction.

The model, delineated in Equation 5.1, utilizes the following parameters to calculate the satisfaction rate (S_{ikt}), which spans from 0% to 100%:

- ε_{ik} : This represents the minimum satisfaction gained by providing the minimum required data rate (β_{min}^k) for content "k".
- β_{min}^k : This is the critical minimum data rate necessary for the application to function at a basic level, below which the application or content "k" becomes unusable.
- β_{max}^k : This is the maximum data rate required to satisfy a user request for content "k".
- D_{ijt}^{down} : This signifies the actual data rate delivered to vehicle "i" from either UAV or RSU "j".

Our model proposes a linear relationship between user satisfaction and the data rates within the bounds defined by the minimum (β_{min}^k) and maximum (β_{max}^k) required data rates for a specific content "k". This approach effectively balances system efficiency and user satisfaction, thereby optimizing overall QoS. The satisfaction rate S_{ikt} is calculated as follows:

$$S_{ikt} = \begin{cases} 0 & D_{ij}^{down} < \beta_{min}^k \\ \varepsilon_{ik} + \frac{100 - \varepsilon_{ik}}{\beta_{max}^k - \beta_{min}^k} \times (D_{ijt}^{down} - \beta_{min}^k) & \beta_{min}^k \leq D_{ij}^{down} \leq \beta_{max}^k \\ 100 & D_{ij}^{down} > \beta_{max}^k \end{cases} \quad (5.1)$$

5.2.3. Traffic Model

In our analysis, we investigate traffic dynamics within both urban settings and highway environments, acknowledging the predominant coverage by RSUs in urban areas versus the reliance on UAVs for highways. Our traffic model accounts for variations in traffic patterns based on time and location, recognizing, for example, the increased traffic volumes in city centers during early morning hours on workdays.

To this end, we adopt the model proposed by Wemegah, Zhu, and Atombo [203], which assesses the impact of different days and road types on peak travel periods. This model is instrumental in generating traffic patterns pertinent to our study, offering insights into the fluctuating nature of vehicular movements.

A critical element of our traffic model is estimating the vehicular density on each road segment. This estimation is crucial for determining the optimal UAV deployment strategy, striking a necessary balance between cost-efficiency and service quality. Such a balance is vital for an adaptive system capable of responding to varied traffic conditions while upholding superior QoS.

5.2.4. UAV Mobility Model

Our UAV mobility model initiates at the operation center, which acts as the launching pad for our UAVs. Upon the need for deployment to a specific location, a UAV is dispatched from the operation center. It ascends to a predetermined altitude, denoted as " h ", before proceeding to the target location at its maximum speed, V_{max} .

The duration of each time slot in our model is defined as δ . Thus, according to equation 5.2, the maximum distance a UAV can traverse within a single time slot is:

$$|loc_j^t - loc_j^{t+1}| \leq V_{max}\delta \quad (5.2)$$

In this equation, loc_j^t and loc_j^{t+1} represent the UAV " j " location at the current time slot " t " and the subsequent time slot " $t + 1$ ", respectively.

Upon arrival at the demand point, the UAV maintains its position to ensure connectivity for users. Essential to this process is the UAV's ability to either directly connect to an RSU or, alternatively, to link with other UAVs that have access to an RSU. Upon reaching the demand point, the UAV hovers at the designated location to ensure connectivity for users. Essential to this process is the UAV's ability to either directly connect to an RSU or, alternatively, to link with other UAVs that have access to an RSU. We define the distance between the UAV and the RSU or vehicles within its coverage area as d_{ij} . This is calculated as follows:

$$d_{ij} = \sqrt{(x_j^{uav} - x_i)^2 + (y_j^{uav} - y_i)^2 + h^2} \quad (5.3)$$

In this equation, x_j^{uav} and y_j^{uav} represent the x and y coordinates of the UAV, respectively, while x_i and y_i represent the x and y coordinates of the RSU or vehicles, respectively. " h " is the altitude of the UAV. This model allows us to account for the spatial dynamics of the UAV's movement and its impact on connectivity.

5.2.5. UAV Energy Consumption Model

Our energy consumption model is largely derived from the empirical studies presented in [204], focusing on the "Intel Aero Ready to Fly Drone", equipped with a 4000mAh battery with a voltage of 14.8V. While energy consumption rates may vary across different UAV models, this specific drone provides a reliable baseline for assessing our method's effectiveness.

The energy consumption model for horizontal movement E_{jt}^{hm} , vertical movement upward E_{jd}^{vu} and hovering E_{jt}^{hover} can be described by the following equations:

$$E_{jt}^{hm} = 308.709\Delta t - 0.852 \quad (5.4)$$

$$E_{jd}^{vu} = 315D - 211.261 \quad (5.5)$$

$$E_{jt}^{hover} = (4.917h + 275.204)\Delta t \quad (5.6)$$

In these equations, Δt represents the duration of flight in the specified situation, " h " denotes the relative hovering altitude, and " D " is the vertical distance traveled upward in meters.

In addition to these factors, UAVs also consume energy during transmission (denoted as E_{jt}^{tr}), which will be elaborated upon in the "Transmission model" section. Therefore, the overall energy consumption model for the UAV can be summarized as:

$$E_{jt}^{tot} = E_{jt}^{tr} + E_{jt}^{hover} + E_{jt}^{hm} + E_{jd}^{vu} \quad (5.7)$$

To estimate the battery life based on energy consumption, we first convert the battery capacity from milliampere-hours (mAh) to ampere-hours (Ah) by dividing by 1000. Next, we convert the battery capacity from Ah to joules, utilizing the following equation.

$$E = Capacity (Ah) * V(Volts) * 3600 \quad (5.8)$$

Given that 1 Joule = 1 Ampere \times 1 Volt \times 1 Second, we can deduce the battery's total energy capacity in joules. Finally, we divide the total energy of the battery in joules by the UAV's average energy consumption in joules per unit time. This results in an estimated battery life, expressed in the same time units as the average energy consumption.

It is important to note that is a simplified estimation. Actual battery life may be affected by various factors, including battery discharge efficiency, power consumption changes, battery age and health, and temperature fluctuations.

5.2.6. Channel Model

The communication channel between UAVs and ground vehicles or RSUs is characterized through a probabilistic path loss model, distinguishing between Line of Sight (LoS) and Non-Line of Sight (NLoS) scenarios.

For LoS scenarios, the average path loss between UAV " j " and vehicle " i " is expressed as:

$$\varphi_{ijt}^{LoS} = 20\log\left(\frac{4\pi f_c d_{ij}}{c}\right) + \eta^{LoS} \quad (5.9)$$

Where, η^{LoS} is system loss for LoS, f_c is the carrier frequency, " c " is the speed of light and d_{ij} denotes the 3-D distance between UAV " j " and the user " i ".

In NLoS scenarios, the average path loss can be defined as:

$$\varphi_{ijt}^{NLoS} = 20\log\left(\frac{4\pi f_c d_{ij}}{c}\right) + \eta^{NLoS} \quad (5.10)$$

Where η^{NLoS} indicates the system loss for NLoS. The expected path loss at time slot " t " is a probabilistic value calculated as follows:

$$\varphi_{ijt} = P_{ijt}^{LoS} \varphi_{ijt}^{LoS} + P_{ijt}^{NLoS} \varphi_{ijt}^{NLoS} = 20n \log \left(\frac{4\pi f_c d_{ij}}{c} \right) + P_{ijt}^{LoS} \eta^{LoS} + P_{ijt}^{NLoS} \eta^{NLoS} \quad (5.11)$$

With P_{ijt}^{LoS} and P_{ijt}^{NLoS} representing the probabilities of having LoS and NLoS conditions, respectively, where $P_{ijt}^{LoS} + P_{ijt}^{NLoS} = 1$.

These probabilities can be calculated by considering different factors such as environmental situations and elevation angles. The probability of having LoS at time slot "t" can be calculated as follows:

$$P_{ijt}^{LoS} = \left(1 + \alpha \exp \left(b \left(\left[\frac{180}{\pi} \right] \theta_{ij} - \alpha \right) \right) \right)^{-1} \quad (5.12)$$

Here, " α ", " b " are environment-dependent constants. For an urban environment, these can be set as 9.6177 and 0.1581, respectively. $[180/\pi]$ is the conversion factor from radians to degrees. With θ_{ij} representing the elevation angle between UAV and vehicle and is calculated as:

$$\theta_{ij} = \arctan(h_j^{uav}/r_{j,i}) \rightarrow r_{j,i} \triangleq \sqrt{(x_j^{uav} - x_i)^2 + (y_j^{uav} - y_i)^2} \quad (5.13)$$

where $r_{j,i}$ is the horizontal distance between UAV "j" and vehicle "i".

The maximum permissible path-loss (φ_{max}) for a successful connection between UAV "j" and vehicle "i" can be defined as follows:

$$\varphi_{max} = p_{U \rightarrow V}^{i,j} + G_U^j + G_V^i - RS \quad (5.14)$$

Here, $p_{U \rightarrow V}^{i,j}$ is the transmitted power, G_U^j is the antenna gain for the UAV "j", G_V^i is the antenna gain for the vehicle "i" and RS is the receiver sensitivity. We would have a successful connection if $\varphi_{ij} < \varphi_{max}$.

The Signal to Interference plus Noise Ratio (SINR) for vehicle "i" when communicating with UAV "j" can now be derived as follows:

$$\gamma_{ijt} = 10 \log p_{U \rightarrow V}^{i,j} - \varphi_{ijt} - 10 \log \left(\sum_{k \neq j} p_{U \rightarrow V}^{i,k} + N \right) \quad (5.15)$$

Where, $p_{U \rightarrow V}^{i,j}$ is the transmission power for the UAV, "N" is the total noise power in the bandwidth of interest. φ_{ijt} is the average path loss, $\sum_{k \neq j} p_{U \rightarrow V}^{i,k}$ is the sum of the powers of the signals transmitted by all other UAVs ($k \neq j$) to vehicle "i". This term represents the total interference from all other UAVs.

This SINR provides a measure of the quality of the signal received by vehicle "i" from UAV "j", taking into account the interference from other UAVs and the background noise level. It's a crucial factor in determining the reliability and quality of the communication link between the UAV and the user.

5.2.7. Transmission Model

In UAV-based wireless communication, the power of the received signal plays a pivotal role in determining the quality and efficiency of data transmission. This power can be modeled using the Friis transmission equation, given certain assumptions like free-space propagation. The received power at time slot "t", expressed as $p_{received}^{i,j,t}$, is calculated as follows:

$$p_{received}^{i,j,t} = p_{U \rightarrow V}^{i,j} + G_U^j + G_V^i - \varphi_{ijt} \quad (5.16)$$

Here, $p_{U \rightarrow V}^{i,j}$ represents the transmitted power, G_U^j is the antenna gain for UAV "j", G_V^i is the antenna gain for the vehicle "i", and φ_{ijt} is the expected path loss.

Once the received power is calculated, it opens up the possibility of determining instantaneous data rates. This is particularly important in designing adaptive communication systems that adjust their transmission parameters based on the prevailing network conditions. The Shannon Capacity formula serves as the key for calculating the downlink data rate at a given time slot "t".

$$D_{ijt}^{down} = W \log_2(1 + \gamma_{ijt}) = W \log_2\left(1 + \frac{p_{received}^{i,j,t}}{N}\right) \quad (5.17)$$

In this equation, "W" is the channel bandwidth, γ_{ijt} is the SINR, $p_{received}^{i,j,t}$ is the received power, and "N" is the noise power (the thermal noise power linearly proportional to the allocated bandwidth and temperature). In the same manner, the data rate for uplink can be calculated.

However, not all data is equal. Different applications and content types necessitate unique data rate requirements, driven by their inherent characteristics and user expectations. For example, video streaming would require a higher data rate compared to a simple file download. To ensure a satisfactory user experience, a minimum data rate is defined. Our model represents this using Eq. 5.18, where the instantaneous data rate to a vehicle from a UAV must exceed a threshold.

$$D_{ijt}^{down} > R_{ikt} \beta_{min}^k \quad \forall k \quad (5.18)$$

Where R_{ikt} is 1 if vehicle "i" requests content "k" at time slot "t" and 0 otherwise, and β_{min}^k is the minimum data rate required to satisfy a user request for content "k".

Moreover, for an efficient operation of UAV-based wireless networks, it is crucial to establish the minimum transmit power required to meet the rate requirement of ground users, as shown in Equation 5.19. Power management is a critical component, affecting system performance, sustainability, and cost-effectiveness:

$$p_{U \rightarrow V, min}^{i,j,t} = \left(2^{D_{ijt}^{down}/W} - 1\right) N \varphi_{ijt} \quad (5.19)$$

Equally essential is the need to quantify the energy consumption for data transmission, which feeds into the system's sustainability and operational cost. In our model, we consider the cumulative transmitted power from a UAV at each time slot as the representative of

energy consumption.

$$E_{jt}^{tr} = \sum_{t=1}^T p_{U \rightarrow V}^{i,j,t} \quad (5.20)$$

From a user's perspective, it is not only about the quality of connection but also how long it takes to receive the requested content. To that end, we introduce a metric estimating the number of timeslots a vehicle requires to receive content from a UAV, as depicted in Equation 5.21. This metric takes into account the content size, the instantaneous data rate, and delays associated with streaming content from neighbor nodes and the cloud. Note that this is an estimate; the actual delay may vary due to dynamic factors such as vehicle movement and fluctuations in the data rate.

$$T_{ijk}^t = \frac{Z_k}{\delta D_{ijt}^{down}} + DL_{neighbor}[C_{jkt} = 0] + DL_{cloud}[C_{jkt} = 0 \text{ for all } j] \quad (5.21)$$

Here, δ is the length of one time slot, Z_k is the content size, D_{ijt}^{down} is the instantaneous data rate to the vehicle "i" from UAV "j", $DL_{neighbor}$ is the delay to stream the content from neighbor nodes (UAV or RSU), DL_{cloud} is the delay to stream the content from the cloud, and C_{jkt} equals 1 if UAV "j" caches the content "k". This equation is based on Iverson bracket notation.

This comprehensive modeling of the Transmission Model, incorporating received power, data rates, transmit power, energy consumption, and estimated delays, forms a robust framework to understand and optimize UAV-vehicle communication dynamics in a complex urban environment.

5.2.8. Blockchain Model

In the context of our study, we utilize blockchain technology to function as a distributed ledger, creating an interconnected network of both UAVs and RSUs. Each UAV and RSU operates as a blockchain node, capable of reading and writing to the blockchain. This system harnesses blockchain not only for its ability to ensure secure, tamper-proof transactions but also to facilitate decentralized operations across the vehicular network, enhancing security and enabling robust decentralization.

The primary motivations for incorporating blockchain into our IoV system are its decentralization capabilities and enhanced security measures. By decentralizing network management, blockchain allows for a more resilient infrastructure, free from the vulnerabilities of centralized control points. Security is intrinsically bolstered through cryptographic mechanisms that safeguard data integrity and authenticate user interactions.

Blockchain plays a critical role in our caching strategy. Each node computes a suitability score for each content item based on characteristics and estimated popularity from request patterns within its coverage area. These scores are then recorded on the blockchain, ensuring that all nodes have access to this predictive data. This setup fosters a coordinated approach to caching without the need for direct communication between nodes. Nodes use this shared information to assess content importance and demand in their vicinity, making informed

caching decisions that align with network-wide priorities. If a node identifies another node as more suitable for caching specific content (due to factors like higher local demand or better storage efficiency) it defers to that node for caching. This decentralized decision-making process effectively reduces redundancy and optimizes storage utilization across the network, ensuring content is cached where it will be most impactful.

For UAV deployment, the prediction of data rate requirements is recorded on the blockchain. A node with adequate processing power and available resources is selected through a consensus mechanism called Proof of Capability (POC) to process this data and make informed decisions about UAV deployment. This approach ensures that decisions regarding UAV positioning and resource allocation are made based on up-to-date, comprehensive network data, enhancing the responsiveness and efficiency of the vehicular network.

In summary, blockchain's integration into our system extends beyond mere data recording; it underpins a sophisticated framework for decentralized, secure decision-making across caching and UAV deployment strategies, significantly enhancing the operational efficacy and security of the IoV environment.

5.3 Proposed Caching Algorithm

Infotainment content, encompassing diverse media such as images, audio, and video, are considered the primary content types of interest within the IoV network. The expeditious delivery of this content types is crucial to enhancing user experience. However, accessing these voluminous data directly from the cloud can lead to increased latency, decreased user satisfaction, and significant strain on the backhaul network. To tackle these challenges, our strategy introduces a sophisticated caching mechanism, crucial for minimizing latency and contributing to a positive user experience by ensuring timely content delivery.

Building upon our previously introduced Federated Learning-based Mobility-aware Collaborative Content Caching (FM3C) strategy [16], we refine the process of content selection for caching. FM3C evaluates content based on a series of metrics, ensuring only the most relevant and demand-driven content is cached. This method, tailored for the limited storage capacities of RSUs and UAVs, dictates a selective caching strategy essential for efficient network management.

To complement the FM3C strategy, we introduce a cooperative caching model. This innovative approach allows nodes not only to respond to local requests but fulfill content requests from adjacent nodes, promoting a collaborative network environment. This cooperation extends beyond simply eliminating redundancy; nodes also collaborate to identify the optimal node for caching specific content, enhancing the overall efficiency of the system.

In traditional setups, such cooperative caching mechanisms would require centralized coordination. However, in our proposed model, we employ blockchain technology, thereby facilitating decentralized cache orchestration. This eliminates the need for a central control unit, facilitating transparent and autonomous decision-making across the network. Integrating FM3C with blockchain for cooperative caching represents a significant leap towards realizing an efficient, scalable content delivery infrastructure within the IoV ecosystem.

5.3.1. FM3C

The FM3C framework represents our tailored solution to manage infotainment content caching in the IoV network. FM3C evaluates several critical factors such as content size, the presence of content replicas, time sensitivity, and, most importantly, content popularity to

optimize caching decisions. By prioritizing popular content for caching, FM3C significantly reduces network latency and enhances user experience by fulfilling a larger volume of requests locally, thus conserving network resources and storage space.

Popularity, however, is not a static attribute but varies over time, necessitating a dynamic approach to predict future content demand accurately. Our solution employs a machine learning model designed to forecast number of content requests. This model processes time-stamped historical request data and content characteristics to anticipate which content will be sought after in subsequent periods.

Given the need to preserve user privacy during this process, we employ federated learning, allowing vehicles to collaboratively contribute to the model's training with their local data without exposing sensitive information. In the first phase of federated learning, each node (RSU or UAV) selects several vehicles within their coverage area to participate in the learning process. This selection is based on the vehicles' capabilities, including their processing power, dataset size, and link quality, among others.

In the next phase, the chosen vehicles download the base model from their respective RSU and begin local training. Once the training phase concludes, the vehicles transmit the parameters of the trained model back to the RSU for aggregation.

Following the aggregation of the models, each node can estimate the popularity of the content within its coverage area over a future time horizon. With popularity as a core factor, each node can calculate a value for each content indicating its relative worth for caching.

However, nodes independently choosing content for caching can result in a considerable waste of storage due to content redundancy. In light of the nodes' ability to cooperate in responding requests, redundant caching should be avoided unless the content is extremely popular or time-sensitive. This cooperative approach maximizes storage efficiency across the network and supports a more effective content delivery system, avoiding the pitfalls of isolated decision-making and redundancy in content storage.

5.3.2. Blockchain-Enabled Caching Strategy

The FM3C framework harnesses blockchain technology to enhance the caching mechanism within the IoV network, shifting away from the need for centralized control to a more agile, decentralized caching strategy. By recording the estimated future popularity of content within specific coverage areas, alongside other content-specific metrics, blockchain enables a nuanced understanding of content demand across the network. This approach allows each node to make informed decisions about what content to cache.

Within this decentralized architecture, blockchain plays a crucial role in facilitating the autonomous decision-making process for caching. After nodes estimate the future popularity of content, these estimates are securely recorded on the blockchain. This ensures all nodes have access to the same predictive data, fostering a coordinated approach to caching without direct communication. Nodes utilize this shared information to assess the relative importance and demand for content in their vicinity, guiding their caching decisions to align with network-wide priorities.

Crucially, the blockchain framework underpins a sophisticated strategy to optimize caching decisions across the network, without direct inter-node communication. Each node calculates the suitability of caching specific content based on a comprehensive analysis that includes its own estimated content popularity within its coverage area and static content-specific data. Since blockchain technology facilitates the sharing of estimated content popularity across different areas, a node can also assess the suitability of caching a content in other nodes'

areas. If a node discerns that another node is better suited to cache a particular content (perhaps due to higher local demand or better storage efficiency) it will defer caching that content, opting instead to let the more suitable node cache it. This decentralized decision-making process efficiently reduces redundancy and enhances the network's overall storage utilization, ensuring that content is cached in locations where it will be most impactful.

The shared ledger on the blockchain provides a transparent record of caching decisions and content popularity metrics, enabling nodes to optimize their caching strategies dynamically. This transparency ensures that all nodes are aware of the content cached by their neighbors, allowing for a more effective cooperative response to user requests. By strategically caching content and leveraging the collective intelligence of the network, nodes can significantly reduce latency and improve the accessibility of high-demand content.

The blockchain not only streamlines the decision-making process for what content to cache but also ensures that the network adapts to changing content demand patterns without the overhead of centralized administration.

5.4 UAV Deployment Strategy

Establishing a comprehensive and reliable infrastructure for the IoV presents significant challenges, particularly in remote or densely populated areas where traditional infrastructure is either economically impractical or infeasible due to location constraints. Furthermore, even in areas where infrastructure does exist, there might be instances such as during high-profile events or festivals where the existing infrastructure might become overwhelmed due to a sudden influx of vehicles.

In response, our research advocates for the strategic use of UAVs to supplement existing RSUs or to provide temporary connectivity where conventional infrastructure is absent. This approach enhances the IoV's connectivity robustness and scalability, offering a flexible solution that transcends the limitations of fixed infrastructure.

The effectiveness and economic viability of UAV deployment hinge on the precise prediction of vehicular traffic. A nuanced understanding of future traffic patterns enables the strategic placement and quantity adjustment of UAVs, ensuring optimal service coverage while efficiently managing deployment costs.

Subsequent subsections will delve into the methodology for estimating vehicular traffic, the formulation of our UAV deployment model based on these estimations, and the application of reinforcement learning to address the complexities of real-time deployment decisions. These details will illustrate the core mechanisms of our proposed UAV deployment strategy, which is designed to strike an optimal balance between cost-effectiveness and service quality.

5.4.1. Traffic and Data Rate Estimation Methodology

In the IoV ecosystem, the strategic deployment of UAVs is intricately linked to the anticipated data rate requirements of users in various areas, which, while influenced by vehicular traffic patterns, fundamentally dictate the number and placement of UAVs. These requirements are subject to fluctuations based on factors like time of day and week, underscoring the necessity for an adaptable prediction model.

To accurately forecast these data rate demands, this study employs LSTM networks, a subset of Recurrent Neural Networks (RNNs) renowned for their capacity to capture long-term dependencies. LSTMs adeptly navigate the challenges of vanishing and exploding gradients,

common in traditional RNNs, through the use of input, output, and forget gates. These mechanisms enable the selective retention and discarding of information, ensuring the preservation of relevant data over extended periods.

The LSTM model is trained on historical traffic data, enriched with contextual information such as day and week, to predict future user data rate requirements. This comprehensive dataset is harvested by RSUs, which continuously monitor vehicular presence and movement within their coverage zones. In areas beyond the reach of RSUs, bordering RSUs at the interface of equipped and unequipped regions collaboratively document vehicle transitions into these zones. Additionally, UAVs present in the area contribute by recording data rate requirements to the blockchain, ensuring a comprehensive dataset is available for analysis.

A node endowed with sufficient computational resources, invariably an RSU due to the processing and power constraints of UAVs, is tasked with executing the LSTM model. The model's output provides a detailed forecast of expected data rate demands across different locations, pinpointing 'demand points' where predicted needs surpass the capacity of existing infrastructure or in regions devoid of RSUs. This predictive insight facilitates a nuanced approach to the strategic placement of UAVs, not only to enhance connectivity but also to address broader IoV system requirements, including data rate provision and network scalability.

5.4.2. Formulation of the Problem

In accordance with the previously established models, our objective is to both maximize user satisfaction rate and minimize the latency experienced by the users. This optimization aims to deploy the minimal possible number of UAVs, while ensuring that each terrestrial user is efficiently serviced by one UAV, and the connectivity of UAVs to the backhaul network is preserved.

To determine if UAV "j" is directly or indirectly connected to an RSU, we define the binary variable R_j as:

$$R_j = \begin{cases} 1 & \tau_j = 1 \text{ or } \sum_{k \in J} \Lambda_{jk} R_k \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad (5.22)$$

τ_j denotes whether UAV "j" has direct access to an RSU. Λ_{jk} equals 1 if UAV "j" is connected to UAV "k", and 0 otherwise.

Given the aforementioned premises, we propose the following Pareto-based multi-objective optimization model:

$$\begin{aligned} & \text{Min} \sum_{j \in J} U_j & (5.23) \\ & \text{Max} \sum_{t \in T} \sum_{k \in K} S_{ikt} \\ & \text{Min} \sum_{t \in T} \sum_{k \in K} T_{ijk}^t \end{aligned}$$

$$\text{Min} \sum_{t \in T} \sum_{i \in I} E_{it}^{tot}$$

S. t:

$$C1: \sum_{j \in J} A_{ij} \geq 1$$

$$C2: D_{ij}^{down} A_{ij} \geq \beta_{min}^k R_{ik} A_{ij} \quad \forall k$$

$$C3: A_{ij} \leq U_j$$

$$C4: D_{ik}^{UAV} \Lambda_{jk} \geq D_{ik,min}^{UAV} \Lambda_{jk}$$

$$C5: \Lambda_{jk} \leq U_{j'}, \Lambda_{jk} \leq U_k$$

$$C6: \sum_{k \in J} R_j \geq 1$$

$$C7: \sum_{k=0}^K C_{jkt} Z_k \leq \psi_j, \quad \forall j, t$$

$$C8: \sum_{t \in T} E_{it}^{tot} < E_i^{Capacity} \quad \forall i$$

Variable Definitions Recap:

To aid comprehension and streamline reference, we provide a consolidated overview of the key variables central to our optimization model. This summary is structured to support reader understanding without necessitating backward navigation through the text:

- **Deployment Status (U_j):** Indicates whether UAV "j" is actively deployed within the operation area. This variable is critical to optimizing UAV deployment for efficient network coverage.
- **User Satisfaction (S_{ikt}):** Reflects the satisfaction level of user "i" accessing content "k" at time "t", highlighting our goal to maximize network service quality.
- **Access Latency (T_{ijk}^t):** Quantifies the delay experienced by user "i" in accessing content "k" from UAV "j", highlighting our aim to minimize network latency.
- **Association Status (A_{ij}):** Specifies the connectivity between user "i" and UAV "j", crucial for understanding network topology and user-UAV interactions.
- **Data Rate Requirements ($\beta_{min}^k, D_{ij}^{down}, D_{ik}^{UAV}, D_{ik,min}^{UAV}$):** Describe the data transmission rates crucial for ensuring satisfactory user experiences and facilitating efficient UAV-to-UAV interactions.
- **Inter-UAV Connectivity (Λ_{jk}):** Denotes the connectivity status between UAV "j" and another UAV "k", illustrating the network of UAV interactions within the system.

These summarized definitions are provided for convenience. For detailed descriptions and the broader context of their applications, readers may refer to Section 5.2 where these variables are introduced comprehensively.

Additional Variable Definitions:

Expanding on the model's formulation, we introduce variables that are pivotal but not previously detailed:

- **Content Caching (C_{jkt}):** Indicates whether content "k" is cached in UAV "j", highlighting strategies for data storage to optimize access and delivery.
- **Content Size (Z_k):** denotes the size of content "k", reflecting the storage capacity needed for caching each piece of content.
- **UAV Capacity ($\psi_j, E_i^{Capacity}$):** Relate to the physical limitations of UAVs, including storage capacity and battery life, which influence strategic deployment and operational longevity.

The formulation of our optimization problem is predicated on a Pareto-based multi-objective approach, balancing several competing objectives to enhance the IoV system's efficiency and user satisfaction. Herein, we delve into the specifics of the objective functions and constraints:

Objective Function:

- Minimize the number of deployed UAVs ($\sum U_j$), targeting cost-effective network expansion.
- Maximize user satisfaction ($\sum S_{ikt}$), ensuring high-quality service delivery.
- Minimize access latency ($\sum T_{ijk}^t$), reducing delays in content delivery.
- Minimize total energy consumption ($\sum E_{it}^{tot}$), promoting sustainable UAV operation.

Constraints:

- **C1-C3** ensure that every user is within reach of at least one UAV or RSU, guaranteeing minimum data rate provision.
- **C4-C6** establish backhaul connectivity for UAVs, ensuring no UAV operates in isolation, disconnected from the network infrastructure.
- **C7** Enforces efficient content caching within the storage limits of both UAVs and RSUs, optimizing data access and ensuring an optimal distribution of content across the network.
- **C8** mandates energy consumption boundaries for UAVs, aligning with their operational capacities.

5.4.3. A Reinforcement Learning Approach to the Solution

The Complexity of the Problem and Rationale for Reinforcement Learning

In light of the discussions surrounding the optimization model for UAV deployment, it becomes evident that we are confronted with a problem of substantial complexity, aptly categorized as a mixed-integer nonlinear problem (MINLP). This classification arises because the model necessitates handling both continuous variables, such as path loss, satisfaction rate, energy consumption, and data rate, and also discrete variables, including deployment status, connectivity, and content caching decisions. Significantly, the designation of MINLP underscores the presence of nonlinear relationships inherent in the model, particularly in the calculations of satisfaction, delay, and energy consumption. These nonlinearities introduce additional layers of complexity, as they often complicate the optimization process, making standard linear solution approaches infeasible. Coupled with the need to balance multiple, potentially conflicting objectives and the uncertainties inherent in dynamic operational environments, the problem presents a formidable challenge that demands an innovative approach for its resolution.

Given these challenges and the dynamic nature of IoV systems, a solution that adeptly navigates these complexities with adaptability and efficiency is paramount. Reinforcement

Learning (RL), with its ability to make sequential decisions in uncertain and dynamic environments, emerges as a fitting solution. Modeling our problem as an MDP facilitates the application of RL by leveraging the Markov property. This property ensures that future states depend only on the current state and not on the sequence of events that preceded it, simplifying decision-making in complex scenarios.

Deep Q-Network Implementation

In our work, we employ Deep Q-Network (DQN), an advanced RL algorithm, to address the UAV deployment optimization. DQN enhances traditional Q-Learning by utilizing deep neural networks to approximate Q-values, which represent the quality of particular actions taken in given states. This approach allows for the handling of high-dimensional state and action spaces, which are prevalent in our problem.

Deep Q-Network Implementation

To ensure that our reinforcement learning approach meets high standards of solution quality and operational efficacy, we benchmark its performance against solutions generated by NSGA-II. NSGA-II, or the Non-dominated Sorting Genetic Algorithm-II, is a well-established method for solving multiobjective optimization problems, known for its ability to find a diverse set of Pareto-optimal solutions efficiently. This algorithm is particularly suitable as a benchmark due to its robustness in handling complex, conflicting objectives, which are characteristic of our UAV deployment model.

In our study, NSGA-II serves as a semi-golden model, providing us with a set of high-quality solutions against which we can measure the efficacy of our DQN model. By comparing the performance of DQN with the Pareto fronts produced by NSGA-II, we gain valuable insights into the relative strengths and areas for improvement in our RL-based approach. This comparative analysis also allows us to determine the training duration required for the DQN to achieve performance levels close to those obtained by NSGA-II, thereby ensuring our model's competitiveness in practical scenarios.

State Space

The state space encompasses predictive cumulative data rate requirements across regions, reflecting the anticipated demand for network resources. This prediction is crucial for pre-deploying UAVs to meet future requirements effectively. Additionally, the current UAV deployment status, residual energy levels, and the cached content catalogue form integral components of the state. These elements capture the system's current operational capacity, resource availability, and content distribution status, providing a comprehensive overview for decision-making.

Action Space

The action space includes all feasible actions that can be executed, ranging from deploying UAVs, adjusting their 3D positions, to recalling them to base. This diversity in actions allows for a dynamic response to varying network demands and operational constraints, aiming to achieve the objectives outlined in the optimization model.

Reward Function

The reward function is designed to guide the DQN towards optimal policy by rewarding desirable outcomes and penalizing suboptimal actions. It ranges from -1 to 1, with -1 indicating the least desired scenario and 1 representing the ideal state. The function is primarily driven by data rate requirements and user coverage, with additional considerations for resource efficiency and operational delay. Specific scenarios include penalizing unnecessary UAV deployment or the presence of isolated UAVs, while rewarding efficient data rate provision and energy optimization.

For instance, maintaining a UAV in its current location when it is needed shortly is more

energy-efficient than redeploying another UAV, factoring in the energy costs associated with movement. Similarly, deploying additional UAVs can improve content caching capacity, indirectly reducing access delays, albeit the primary focus remains on ensuring coverage and adequate data rates.

Toward an Optimized Reward Function

Considering the multifaceted objectives of minimizing delay, maximizing data rate satisfaction, and optimizing energy usage, the formulation of the reward function is complex yet deliberate. Our approach systematically weights different aspects to align with their impact on overall network performance. Despite the intricate dynamics among these factors, our methodology incorporates a structured equation that adeptly balances these considerations. The principle behind our formulation is as follows:

- Assigning higher rewards for actions that directly contribute to meeting or exceeding data rate requirements efficiently.
- Penalizing actions leading to resource wastage or suboptimal utilization, such as unnecessary UAV deployments or failing to meet user demands.
- Incrementally adjusting rewards based on the comparative analysis of energy consumption and delay reduction benefits, ensuring a balanced approach to resource management and service quality enhancement.

This nuanced reward structure encourages strategies that not only fulfill immediate network demands but also consider long-term efficiency and sustainability.

5.5 Performance Evaluation

This section presents the findings and results of our comprehensive study, focusing primarily on evaluating the performance of the SONA methodology in the IoV environment. While FM3C strategy remains a foundational component of our caching mechanisms, drawing from our prior work, this evaluation centers on SONA's innovative approach to UAV deployment. The essence of this chapter is to underscore the efficacy of SONA as a significant advancement in enhancing IoV systems, showcasing its capacity to navigate the complexities of UAV deployment and connectivity.

5.5.1. Simulation Environment Description

Our simulation environment is meticulously designed to mirror a realistic IoV landscape, encompassing four urban areas equipped with RSUs and a highway area (alternatively referred to as a rural or remote area) lacking traditional infrastructure, designated for the exclusive reliance on UAVs for connectivity and coverage.

The simulated IoV environment hosts 610 vehicles, navigating across different regions and requesting a variety of contents. Our content catalogue is extensive, featuring 9,742 items, with the real-world MovieLens dataset serving as the basis for simulating content requests. The temporal details of movie ratings within this dataset are ingeniously repurposed to correspond to content request timings in our simulation, adding a layer of realism to our scenario.

A fleet of 32 UAVs is considered available for potential deployment. In alignment with the strategic objectives of SONA, this fleet is not limited to providing coverage in areas devoid of RSUs. It also holds the potential for deployment within urban areas already equipped with RSUs. This strategic flexibility allows for the enhancement of network capacity and QoS in

dense urban regions, while also extending essential connectivity services to remote or rural areas traditionally underserved by existing infrastructure. This dual capability underscores the adaptive and comprehensive nature of the SONA methodology in addressing a wide spectrum of connectivity challenges within the IoV landscape.

5.5.2. Simulation Parameters

For a rigorous evaluation, we've meticulously chosen simulation parameters that reflect realistic operational conditions within the Mid-Band 5G spectrum, ensuring the relevance and applicability of our findings. Table 5.1 summarizing the key parameters utilized in our simulations.

These parameters were selected to offer a balance between realism and computational feasibility, ensuring that our simulations faithfully replicate potential real-world IoV scenarios.

To evaluate the effectiveness of the SONA methodology, its performance is compared against three distinct methodologies developed for this analysis, each serving a unique purpose within our comparative framework.

1. **Oracle Methodology:** This hypothetical model possesses foresight into future events, including precise content request patterns, vehicle placements, and data rate requirements at future timesteps. By virtue of this omniscience, the Oracle method can optimize caching decisions and UAV deployment with an unmatched level of precision. Although its assumptions are not grounded in practical reality, serving as an unattainable upper benchmark, it offers a valuable perspective by setting an ideal standard. The closer a realistic algorithm approaches the Oracle's performance, the more efficient it is considered.
2. **Last Recent Step (LRS):** The LRS method capitalizes on the immediacy of current data, utilizing known variables such as vehicle placements, data rate requirements, and content requests at the present timestep to inform caching and UAV deployment strategies for the imminent future. Operating under the assumption of minimal changes between consecutive timesteps, LRS seeks to pragmatically adjust the network's response to evolving demands, offering a balance between foresight and adaptability.
3. **Random Methodology:** This approach adopts a stochastic strategy for both UAV deployment and content caching. A random selection process dictates the deployment of UAVs and the caching of content, introducing an element of unpredictability.

Parameter	Value	Unit	Notes
Transmission Power	40	dBm (10 W)	Standard power level for UAV communications.
Noise Power	-100	dBm/Hz	A common value for outdoor communication scenarios.
Frequency	6	GHz	Falls within the Mid-Band 5G, suitable for IoV applications.
System Loss (LoS)	2	dB	Typical loss in Line-of-Sight conditions.
System Loss (NLoS)	15	dB	Reflects higher loss in Non-Line-of-Sight conditions.
Antenna Gain	3	dB	Realistic gain for UAV deployed antennas.
Bandwidth	100	MHz	Provides ample capacity for IoV communication needs.

Table 5.1. Summary of simulation parameters.

5.5.3. Simulation Results

The cache hit rate is a pivotal metric in evaluating the performance of content caching strategies within an IoV system. It measures the proportion of content requests that are served from cached data relative to the total number of requests, highlighting the efficiency of content caching. This metric is central to understanding how well caching mechanisms are performing, impacting not only network latency but also other critical aspects like backhaul load reduction and data availability. These factors collectively contribute to the overall effectiveness and responsiveness of an IoV system.

In Figure 5.2, the cache hit rate across different methodologies is compared as storage capacity increases, a variable crucial for understanding caching performance under various conditions. For this analysis, storage size variations are uniformly applied to RSUs in each simulation run, with UAV storage set to 30% of RSU capacity to reflect their relatively lower storage capabilities. This approach ensures a realistic representation of network infrastructure capabilities.

SONA exhibits a performance closely resembling that of the Oracle, outstripping both LRS and Random approaches, particularly at constrained storage capacities. This indicates SONA's proficiency in optimizing cache placement even with limited resources, closely approaching an ideal performance benchmark set by the Oracle. The linear improvement in cache hit rate with increased storage capacity observed for the Random method contrasts starkly with the underperformance of LRS in the dynamic IoV environment. Despite the assumption that LRS could offer advantages by leveraging current data for near-future optimizations, its strategy falls short in the rapidly evolving IoV context, where vehicular movement and demand fluctuate significantly.

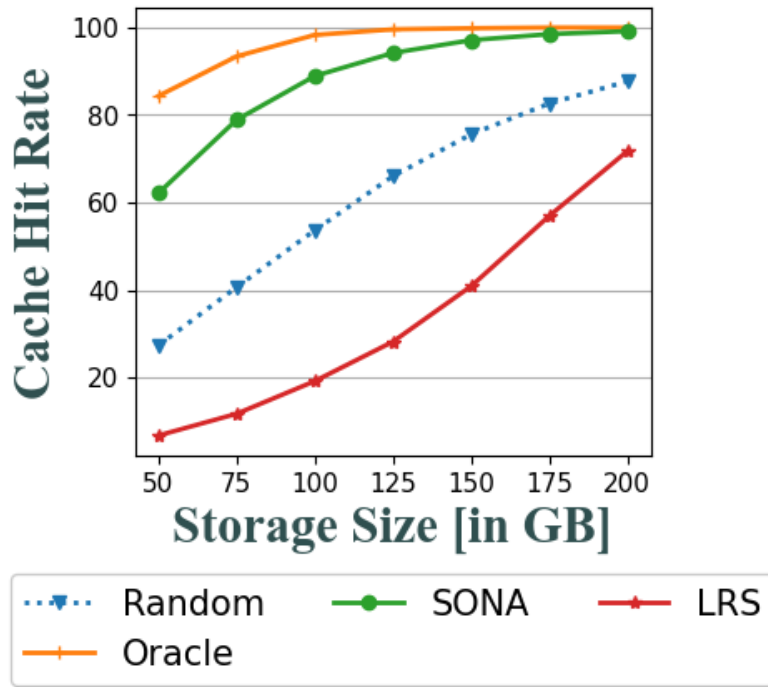


Figure 5.2. Cache Hit Rate Comparison Across Different Storage Sizes for SONA, Oracle, LRS, and Random Methods in the IoV Environment.

Fig. 5.3. reveals key insights into the distribution of cache hit rates between RSUs and UAVs under the SONA framework. A primary observation from this analysis is the significantly larger share of cache hit rates contributed by RSUs compared to UAVs. This outcome aligns with expectations, given RSUs' enhanced capacity for storing a more extensive array of contents, coupled with their stable presence within the network infrastructure. It is important to reiterate that the principal objective of UAV deployment within the SONA methodology is not primarily focused on content caching but rather on augmenting network coverage and connectivity.

Despite the relatively modest local cache hit rates attributed to UAVs, their strategic deployment ensures that users connected to UAVs are not isolated from the broader network's caching capabilities. Each UAV, mandated to maintain either direct or indirect connectivity to the RSU infrastructure, acts as a bridge for users to access cached contents across the network. Consequently, even though direct cache hits from UAVs may be fewer, users connected through UAVs benefit from reduced latency in content delivery. This access to cached content, albeit not locally but through the networked infrastructure, ensures significantly lower delays compared to fetching content directly from the cloud. The specifics of how these dynamics influence overall network delay will be further explored in the

forthcoming discussion on latency metrics in Fig. 5.4.

Figure 5.4. shifts the focus to the impact of these caching strategies on network latency. SONA demonstrates remarkable proficiency in maintaining the average delay at approximately 8 ms, a benchmark conducive to supporting a wide array of IoV applications,

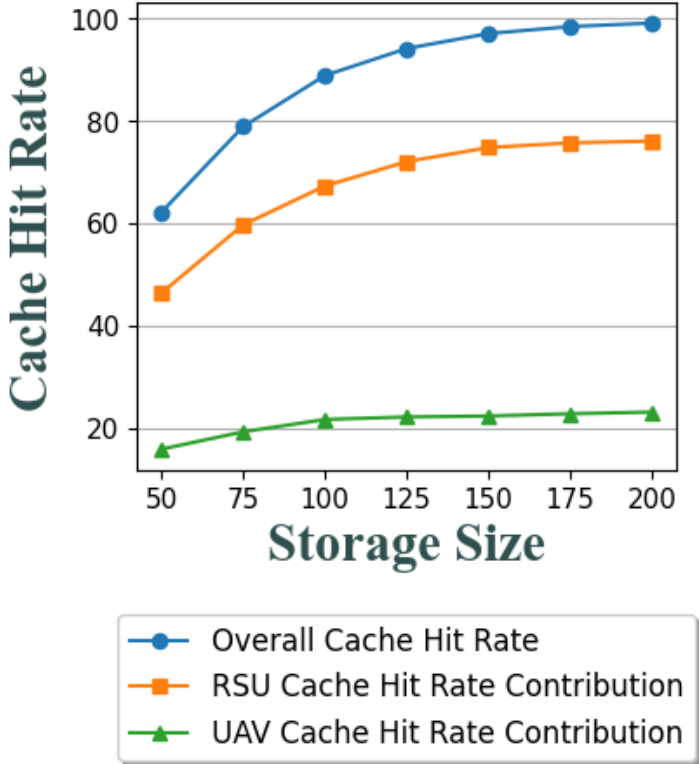


Figure 5.3. Comparative Analysis of Cache Hit Rate Contributions by RSUs and UAVs within the SONA Framework.

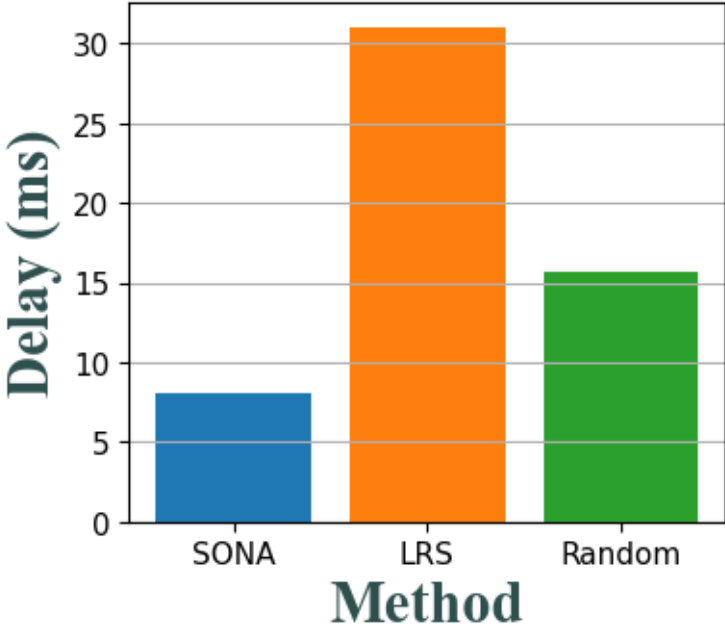


Figure 5.4. Comparison of Average Network Delay Across SONA, Random, and LRS Methodologies.

including those demanding near-real-time responsiveness. This performance starkly contrasts with the Random methodology, which subjects users to an average delay nearly twice that encountered under SONA. More strikingly, LRS trails significantly, resulting in delays that are double those observed with the Random approach.

Figure 5.5. offers a comprehensive look at the average number of UAVs deployed by each method (SONA, Oracle, and LRS) to achieve varying target satisfaction rates ranging from 40% to 90%. Notably, the Random method is absent from this analysis due to its stochastic nature, which does not tailor UAV deployment based on specific satisfaction targets. The average deployment of 15.18 UAVs by the Random method, deduced from its indiscriminate strategy, stands out as significantly higher than the deployments noted for the other methods in this comparison, underscoring a lack of efficiency in UAV utilization.

The data presented in Fig. 5.5 reveals a striking similarity in the average number of UAVs deployed by SONA, Oracle, and LRS to meet the set satisfaction thresholds. This closeness in

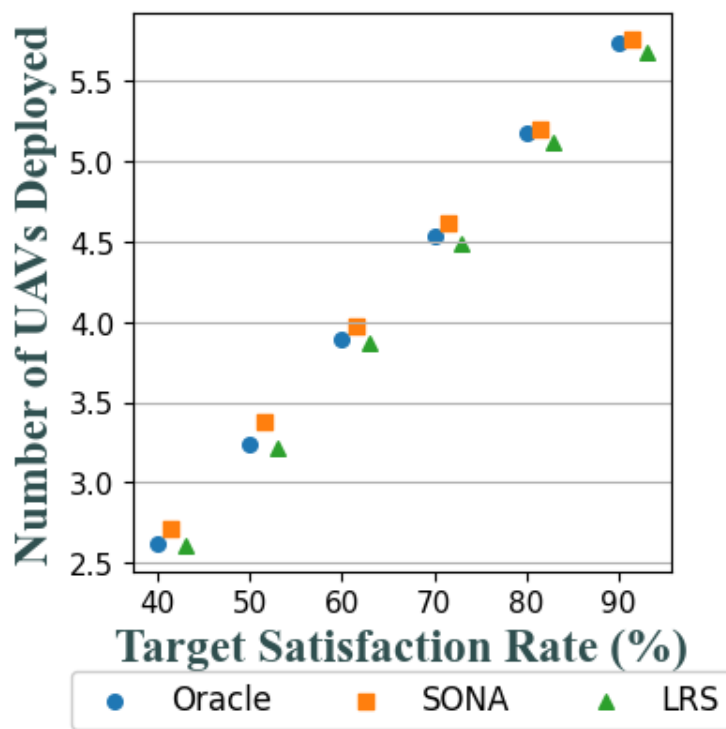


Figure 5.5. UAV Deployment Efficiency to Achieve Target Satisfaction Rates Across Oracle, SONA, and LRS Methodologies.

values, however, stems from distinct operational dynamics inherent to each method. SONA's alignment with Oracle underscores its deployment efficiency, indicating that SONA strategically deploys UAVs in a manner nearly as optimized as the theoretical best-case scenario represented by Oracle. This efficiency highlights SONA's capability to judiciously utilize UAV resources to achieve desired network satisfaction levels, mirroring Oracle's foresight without having access to future data.

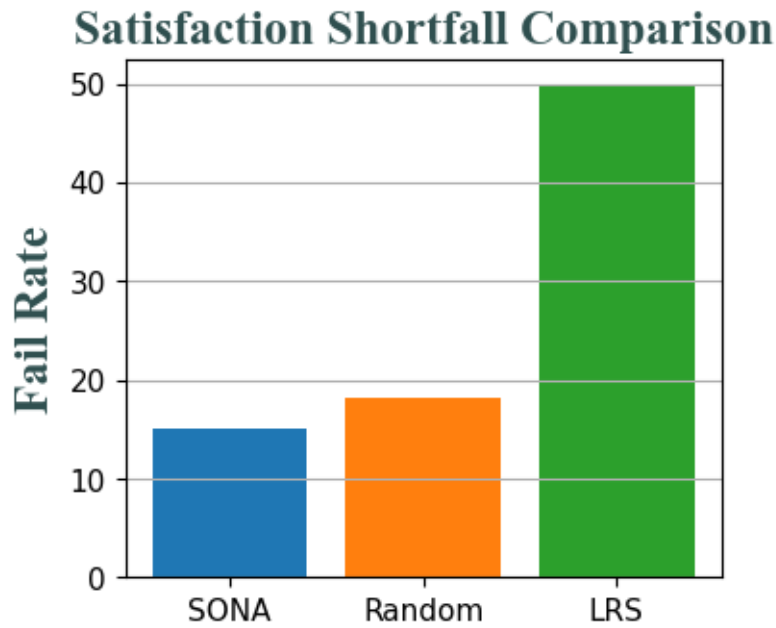


Figure 5.6. Comparative Analysis of Missed Target Satisfaction Timeslots Across Methods.

Conversely, the proximity of LRS's average UAV deployment to Oracle and SONA is attributable to a different rationale. LRS bases its deployment decisions on current timestep data, essentially introducing a one-time-slot lag in its responsiveness. While this approach enables LRS to deploy a number of UAVs closely matching the more foresighted methods, it's the timing, rather than the quantity, where LRS diverges from optimal efficiency. This nuanced distinction is crucial, as it reveals LRS's potential limitations in dynamic environments where network conditions and user demands can shift rapidly from one timestep to the next.

Delving into the nuances of how each method copes with the challenge of meeting a set satisfaction rate, Figure 5.6 casts a spotlight on the frequency of shortfalls from a 70% target satisfaction rate across SONA, Random, and LRS methodologies. This examination shifts our focus from the sheer average number of UAV deployments to the effectiveness of these deployments in consistently achieving desired network satisfaction levels.

In an intriguing revelation, SONA emerges with the lowest failure rate, underscoring its precision in UAV deployment and operational efficiency. This finding is particularly noteworthy in light of the Random method's strategy, which, despite deploying an average of more than three times the number of UAVs compared to SONA, still incurs a higher rate of failure in attaining the target satisfaction rate. This discrepancy accentuates the Random method's inefficiency, where the abundance of resources does not translate to higher success rates, indicating a misalignment between UAV deployment and actual network needs.

More strikingly, LRS exhibits a failure rate exceeding 50%, despite deploying a comparable number of UAVs to SONA, as highlighted in Fig. 5.5. This outcome suggests that while LRS might match SONA in the quantity of deployed UAVs, it significantly lags in deploying them in a manner that consistently meets the target satisfaction rate. The data suggests that LRS's reliance on immediate past data to inform future deployments is insufficient for the dynamic and fluctuating demands of the IoV environment, leading to its higher shortfall frequency. These observations underscore the critical importance of strategic UAV deployment beyond mere numbers.

Building on the insights gained from Fig. 5.6, Fig. 5.7. delves deeper into the nature of each methodology's failures to meet the 70% satisfaction target, providing a clearer picture of the

severity of these shortfalls. This figure quantifies the average gap between the achieved satisfaction rates and the target, offering a lens through which to assess not just the frequency but also the magnitude of failures across SONA, Random, and LRS.

This analysis reveals a compelling aspect of SONA's performance. Not only does SONA exhibit the lowest failure rate in achieving the target satisfaction rate, as shown previously, but it also minimizes the impact when failures do occur. The average shortfall for SONA is notably less than 30%, indicating that even in scenarios where SONA does not fully meet the targeted level of user satisfaction, it still maintains a relatively high-performance level, often remaining within a margin that could be considered acceptable for many applications.

In stark contrast, the Random and LRS methods not only fail more frequently but also with greater severity. The average shortfall for Random approaches 50%, and for LRS, it exceeds this threshold, suggesting that when these methods fail to achieve the target satisfaction rate, the resultant level of service significantly deviates from user expectations. This can render the network service perceived as inadequate or even unusable, highlighting a critical limitation in their ability to maintain acceptable service quality levels.

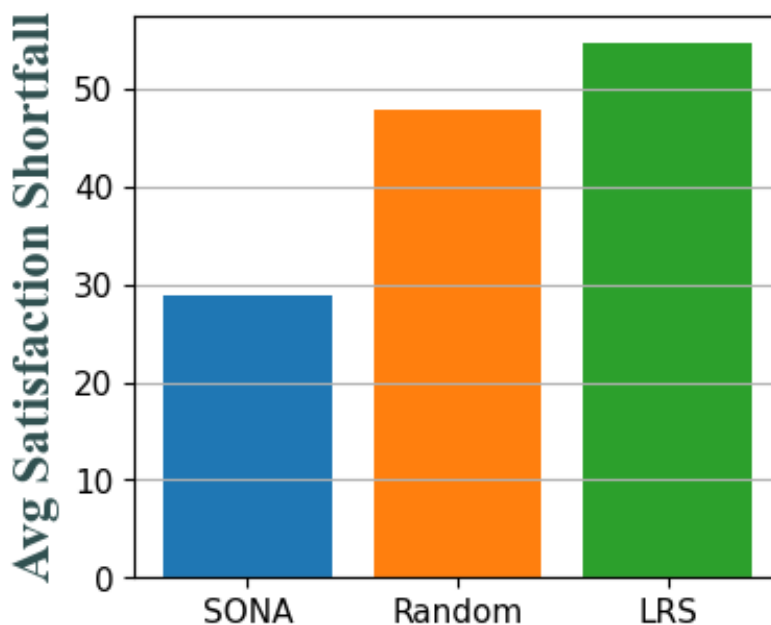


Figure 5.7. Analysis of Satisfaction Shortfall Severity Across Methods.

To effectively evaluate our SONA method, which implements DQN for UAV deployment, we utilize NSGA-II as a semi-golden standard. Known for its efficiency in multiobjective optimization, NSGA-II provides Pareto-optimal solutions that serve as benchmarks for assessing the outputs of our reinforcement learning model.

We calculated the loss for both NSGA-II and the SONA method using Mean Squared Error (MSE) by comparing their results with those from the Oracle method, the ideal performance benchmark in our study. The specific loss value of 0.0068 for NSGA-II, which was used as the stopping threshold, illustrates the deviation of its results from the Oracle's using this metric. This MSE approach was consistently applied to evaluate the SONA method as well.

Fig. 5.8 shows the trajectory of loss values during SONA's training over different epochs, highlighting the point at which the loss meets the 0.0068 threshold, established based on NSGA-II's performance relative to the Oracle. This threshold guides the duration of training, ensuring that our SONA method achieves a competitive level of performance.

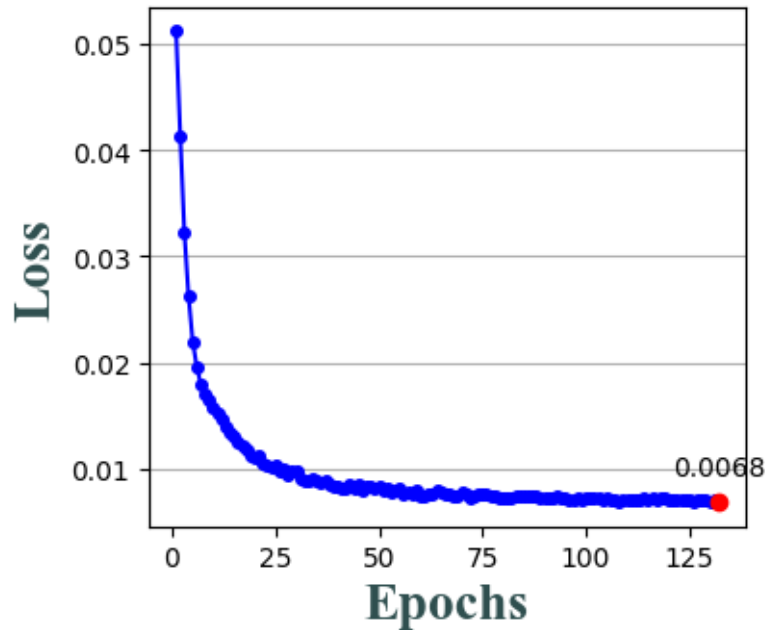


Figure 5.8. Execution Time Comparison Between Reinforcement Learning and

This benchmarking strategy demonstrates that SONA can match the results of NSGA-II after 130 epochs of training, but significantly, it can achieve very similar outcomes after just 25 epochs. This not only underscores the efficiency and accuracy of SONA but also highlights its practical advantages. Once trained, the DQN used in SONA can deliver immediate results for new states, providing on-the-spot decisions without the need for further computations. In contrast, NSGA-II lacks this capability, requiring time-consuming computations for each new state. This difference is critical in dynamic environments where timely decision-making is crucial.

5.6 Conclusion

Reflecting on our journey through the integration of UAV deployment and blockchain technology in the IoV, this study has solidified SONA's role in significantly advancing network services. Through the dynamic positioning of UAVs and the decentralized, secure caching mechanisms enabled by blockchain, we've demonstrated marked improvements in reducing latency and ensuring robust data rate provision. Our exploration underscores the seamless synergy between SONA and the previously developed FM3C strategy, affirming their collective potential in enhancing the IoV landscape. The findings emphasize not only the feasibility but also the efficiency of leveraging these advanced technologies to meet the evolving demands of intelligent transportation systems.

Looking toward the future, the continuing evolution of IoV beckons further exploration into the integration of next-generation technologies and the expansion of IoV applications across various sectors. The exploration of enhanced security measures to safeguard this increasingly complex network against emerging threats will be crucial. As this study lays the groundwork for future innovations, the potential for leveraging SONA within broader IoV applications promises a horizon brimming with opportunities for creating more connected, efficient, and resilient environments.

6. Conclusion and Future Directions

In this thesis, we explored caching techniques within the Internet of Things (IoT) and the Internet of Vehicles (IoV), seeking to improve network performance and user experience in these highly connected environments. The research was divided into three main sections, each addressing specific challenges and proposing new solutions to optimize the flow of data and overall network functionality.

In the first part of this research, presented in Chapter Three, we addressed the challenges associated with caching and routing in the rapidly evolving IoT landscape. The primary goal was to enhance the efficiency of data management and network resource utilization across IoT systems, which are increasingly burdened by the high demands of data traffic and diverse application needs.

The specific problems we aimed to solve involved overcoming the limitations of conventional caching and routing methods, which often struggle to cope with the dynamic nature of IoT environments. Traditional methods do not adequately consider the varying capabilities of IoT devices or the specific requirements of different data types, which can lead to inefficiencies and increased latency.

To address these issues, we introduced several innovative solutions. We implemented a multi-criteria decision-making framework that allowed for smarter caching decisions by evaluating multiple factors, such as data popularity, node storage capabilities, and the network's current state. This approach ensured that data was stored closer to where it was most likely to be accessed, thus reducing retrieval times and network congestion.

Furthermore, we leveraged Software-Defined Networking (SDN) to enable more intelligent routing decisions. By centralizing the control of network flows, SDN provided the flexibility needed to adjust routes dynamically based on current network conditions and data requirements. This was particularly useful in managing the complex interactions and data flows within IoT networks, where device capabilities and connectivity can vary widely.

Additionally, our system incorporated hierarchical network orchestrations to manage data across different layers of the network architecture effectively. This setup allowed for a more organized and scalable approach to network management, suitable for the extensive and multi-tiered nature of IoT systems.

In terms of performance, the solutions we developed led to significant improvements in network efficiency. We achieved up to an 89% cache hit rate, indicating that most requests for data were successfully served from the cache without needing to retrieve data from the main servers, thereby reducing latency and network strain. Moreover, our robust network design ensured that about 60% of network nodes remained functional during 90% of the simulation period, highlighting the reliability and resilience of the system under typical operational conditions.

These results underscored the effectiveness of the proposed multi-criteria decision-making approach and the strategic use of SDN in enhancing the performance and reliability of IoT networks, setting a strong foundation for future advancements in IoT network management.

Moving on to Chapter Four, we tackled the distinct challenges of caching within the IoV. Our primary objective was to refine content delivery systems in IoV by tailoring caching mechanisms to suit various types of content, specifically addressing the unique demands of safety-related and infotainment data. The need for such differentiation stems from the critical nature of timely and reliable data delivery in vehicular environments, where the quality of service can directly impact user safety and experience.

We introduced proprietary caching solutions designed to optimize the storage and retrieval of both emergency (safety) and non-emergency (infotainment) content. For safety applications, our caching strategy prioritized rapid access and local relevance, ensuring that data pertaining to safety could be swiftly and reliably accessed by vehicles in the vicinity of relevant incidents. For infotainment content, the focus was on minimizing latency, considering the large size of multimedia files and their less critical nature.

A significant innovation in this work was the implementation of a privacy-preserving mechanism using federated learning for popularity prediction. This approach allowed us to predict the demand for specific content types without compromising the privacy of individual users. By leveraging data from multiple vehicles in a collaborative yet decentralized manner, we were able to gather accurate insights into content popularity without directly accessing sensitive user data.

Moreover, we developed a new candidate selection mechanism and aggregation method for the federated learning process. This enhanced the efficiency and accuracy of our federated models by selectively incorporating data from nodes that provided the most relevant and significant insights, thus refining the learning process and outcomes.

The results from these innovations were compelling. We achieved up to a 98% cache hit rate on a medium storage size, which significantly reduced the need for vehicles to fetch content from distant servers, thereby minimizing latency. Indeed, we managed to reduce the delay to under 10ms for the same storage configuration. Additionally, the simulation results demonstrated the superiority of our proposed candidate selection and aggregation methods in the federated learning setup, which were key to enhancing the overall performance and efficiency of the IoV caching system. These outcomes not only highlight the effectiveness of tailored caching strategies but also underscore the potential of advanced machine learning techniques in managing the complex dynamics of vehicular networks.

In the third and final segment of this thesis presented in Chapter Five, we addressed the complex challenges of network coverage and data accessibility in the IoV by focusing on the strategic deployment of Unmanned Aerial Vehicles (UAVs). The primary goal was to enhance the network's reach and reliability, particularly in areas underserved by traditional road-side units (RSUs) or where temporary coverage boosts are necessary, such as in crowded events or disaster response scenarios.

We introduced an innovative UAV deployment strategy named SONA, designed to strategically position UAVs to not only assist in data caching but also significantly expand network coverage. This approach was underpinned by the use of blockchain technology for decentralized network orchestration, ensuring a secure and transparent method for managing the dynamic and distributed nature of UAV operations. Blockchain's role was critical in maintaining a consistent and tamper-proof ledger of UAV positions, movement decisions, and data caching statuses, facilitating cooperative decision-making across the network.

To optimize the deployment of UAVs, we employed a Pareto Optimization framework that balanced multiple objectives, such as minimizing deployment costs while maximizing coverage and data accessibility. This multi-objective approach allowed us to evaluate trade-offs effectively and make decisions that align with the overarching goals of network efficiency and user satisfaction.

Furthermore, we transformed this optimization challenge into a Markov Decision Process (MDP) to dynamically analyze UAV movement and content request patterns. This transformation was crucial for applying Deep Reinforcement Learning (DRL), which provided a robust solution framework capable of adapting to changing network conditions and user demands in real-time. DRL was particularly adept at navigating the complexities of UAV

positioning, helping to ensure optimal coverage and service levels across various scenarios.

The results of implementing SONA were highly promising. Our method achieved an average delay of less than 8 milliseconds, which is critical for supporting real-time communication and high-quality service delivery within IoV systems. Additionally, SONA managed to maintain higher satisfaction rates with fewer UAVs deployed compared to traditional methods. This efficiency in UAV usage not only reduced operational costs but also lessened the environmental impact, aligning with sustainable deployment practices. The strategic deployment of UAVs, guided by advanced optimization and machine learning techniques, highlighted our ability to significantly enhance the adaptability and responsiveness of IoV infrastructures.

The proposed solutions not only make these networks more efficient but also open up new possibilities for future improvements, especially with the advent of technologies like 5G and AI. This work not only pushes the envelope in network management but also sets the stage for further research and development in intelligent transportation and interconnected devices, aiming for systems that are more responsive, efficient, and robust.

Looking forward, several avenues of research emerge from the foundations laid by this thesis, particularly in enhancing the robustness and user engagement of the proposed mechanisms. One critical area to explore is the security aspect of the systems developed, especially given their decentralized nature and reliance on cooperative network behavior. While blockchain inherently provides a layer of security through its cryptographic functions, the specific security challenges of UAV communications, federated learning, and decentralized caching have not been fully addressed. Investigating potential vulnerabilities and developing robust security protocols could significantly enhance the trustworthiness and reliability of these systems. For example, securing the communication channels between UAVs and ground nodes, protecting the integrity of data in federated learning, and safeguarding against malicious attacks in decentralized caching are essential for the widespread adoption of these technologies.

Another compelling direction for future research lies in the design of incentive mechanisms to encourage user participation, which is crucial for the success of systems like federated learning where user cooperation directly impacts the effectiveness of the model. Given the decentralized setups of the mechanisms proposed in this thesis, creating economic or data-driven incentives could promote more active participation and richer data contributions from users. This not only improves the accuracy and efficiency of the models but also potentially enriches the user experience by making the systems more responsive and tailored to real-time needs. Exploring different models of incentives, from micropayments to improved service offerings for contributing users, could provide the necessary motivation for users to engage more deeply with these IoV technologies. These areas of research would not only address key challenges but also significantly advance the field by enhancing the scalability, security, and user-centric aspects of IoT and IoV ecosystems.

Bibliography

- [1] N. Sharma, M. Shamkuwar, and I. Singh, "The history, present and future with IoT," *Internet of things and big data analytics for smart generation*, pp. 27–51, 2019.
- [2] M. Wollschlaeger, T. Sauter and J. Jasperneite, "The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0," in *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, pp. 17-27, March 2017
- [3] C. Paul, A. Ganesh and C. Sunitha, "An overview of IoT based smart homes," 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2018, pp. 43-46
- [4] S. Selvaraj and S. Sundaravaradhan, "Challenges and opportunities in IoT healthcare systems: a systematic review," *SN Applied Sciences*, vol. 2, no. 1, p. 139, 2020.
- [5] J. Xu, B. Gu, and G. Tian, "Review of agricultural IoT technology," *Artificial Intelligence in Agriculture*, vol. 6, pp. 10–22, 2022.
- [6] S. Muthuramalingam, A. Bharathi, S. Rakesh Kumar, N. Gayathri, R. Sathiyaraj, and B. Balamurugan, "IoT based intelligent transportation system (IoT-ITS) for global perspective: A case study," *Internet of Things and Big Data Analytics for Smart Generation*, pp. 279–300, 2019.
- [7] Z. Mahmood, "Connected vehicles in the IoV: Concepts, technologies and architectures," *Connected vehicles in the internet of things: concepts, technologies and frameworks for the IoV*, pp. 3–18, 2020.
- [8] A. Thakur and R. Malekian, "Internet of vehicles communication technologies for traffic management and road safety applications," *Wireless Personal Communications*, vol. 109, no. 1, pp. 31–49, 2019.
- [9] M. Bansal, I. Chana, and S. Clarke, "A survey on iot big data: current status, 13 v's challenges, and future directions," *ACM Computing Surveys (CSUR)*, vol. 53, no. 6, pp. 1–59, 2020.
- [10] Y. Liu, H.-N. Dai, Q. Wang, M. K. Shukla, and M. Imran, "Unmanned aerial vehicle for internet of everything: Opportunities and challenges," *Computer communications*, vol. 155, pp. 66–83, 2020.
- [11] F. Pereira, R. Correia, P. Pinho, S. I. Lopes, and N. B. Carvalho, "Challenges in resource-constrained IoT devices: Energy and communication as critical success factors for future IoT deployment," *Sensors*, vol. 20, no. 22, p. 6420, 2020.
- [12] S. A. Hussain, K. M. Yusof, S. M. Hussain and A. V. Singh, "A Review of Quality of Service Issues in Internet of Vehicles (IoV)," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 380-383
- [13] S. Vural, P. Navaratnam, N. Wang, C. Wang, L. Dong and R. Tafazolli, "In-network caching of Internet-of-Thing's data," 2014 IEEE International Conference on Communications (ICC), Sydney, NSW, Australia, 2014, pp. 3185-3190
- [14] E. K. Zavadskas, Z. Turskis, and S. Kildienė, "State of art surveys of overviews on MCDM/MADM methods," *Technological and economic development of economy*, vol. 20, no. 1, pp. 165–179, 2014.
- [15] S. Khodaparvas, A. Benslimane, and S. Yousefi, "A software-defined caching scheme for the Internet of Things," *Computer Communications*, vol. 158, pp. 178–188, May 2020,
- [16] S. Khodaparvas, A. Benslimane, and S. Yousefi, "An Intelligent Caching Scheme Considering the Spatio-Temporal Characteristics of Data in Internet of Vehicles," *IEEE Trans. Veh. Technol.*, pp. 1–15, 2023,
- [17] S. Khodaparvas, S. Yousefi, and A. Benslimane, "A Multi Criteria Cooperative Caching Scheme for Internet of Things," in *ICC 2019 - 2019 IEEE International Conference on Communications*

- (ICC), Shanghai, China: IEEE, May 2019, pp. 1–6.
- [18] B. B. Gupta and M. Quamara, “An overview of Internet of Things (IoT): Architectural aspects, challenges, and protocols,” *Concurrency and Computation*, vol. 32, no. 21, p. e4946, Nov. 2020,
- [19] ITU-T Y.2060. <https://handle.itu.int/11.1002/1000/11559>, Accessed 15 December 2023.
- [20] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of Things (IoT): A vision, architectural elements, and future directions,” *Future generation computer systems*, vol. 29, no. 7, pp. 1645–1660, 2013.
- [21] K. K. Patel, S. M. Patel, and P. Scholar, “Internet of things-IOT: definition, characteristics, architecture, enabling technologies, application & future challenges,” *International journal of engineering science and computing*, vol. 6, no. 5, 2016.
- [22] L. D. Xu, E. L. Xu, and L. Li, “Industry 4.0: state of the art and future trends,” *International Journal of Production Research*, vol. 56, no. 8, pp. 2941–2962, Apr. 2018,
- [23] S. Madakam, R. Ramaswamy, and S. Tripathi, “Internet of Things (IoT): A literature review,” *Journal of Computer and Communications*, vol. 3, no. 5, pp. 164–173, 2015.
- [24] Cisco Annual Internet Report (2018–2023) White Paper. <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>, Accessed 10 March 2024.
- [25] R. R. Chowdhury, A. C. Idris, and P. E. Abas, “Identifying SH-IoT devices from network traffic characteristics using random forest classifier,” *Wireless Networks*, vol. 30, no. 1, pp. 405–419, 2024.
- [26] S. C. Mukhopadhyay, “Wearable sensors for human activity monitoring: A review,” *IEEE sensors journal*, vol. 15, no. 3, pp. 1321–1330, 2014.
- [27] J. Klonovs et al., *Distributed computing and monitoring technologies for older patients*. Springer, 2015.
- [28] E. I. Konstantinidis, P. E. Antoniou, G. Bamparopoulos, and P. D. Bamidis, “A lightweight framework for transparent cross platform communication of controller data in ambient assisted living environments,” *Information Sciences*, vol. 300, pp. 124–139, 2015.
- [29] Y. E. Aslan, I. Korpeoglu, and Ö. Ulusoy, “A framework for use of wireless sensor networks in forest fire detection and monitoring,” *Computers, Environment and Urban Systems*, vol. 36, no. 6, pp. 614–625, 2012.
- [30] R. Tan, G. Xing, J. Chen, W.-Z. Song, and R. Huang, “Fusion-based volcanic earthquake detection and timing in wireless sensor networks,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 9, no. 2, pp. 1–25, 2013.
- [31] S. R. Nimbargi, S. Hadawale, and G. Ghodke, “Tsunami alert & detection system using IoT: A survey,” in *2017 International Conference on Big Data, IoT and Data Science (BIG DATA, IoT and DS)*, IEEE, 2017, pp. 182–184.
- [32] A. Van Herwijnen and J. Schweizer, “Monitoring avalanche activity using a seismic sensor,” *Cold Regions Science and Technology*, vol. 69, no. 2–3, pp. 165–176, 2011.
- [33] J. Cynthia, M. Saroja, P. Sultana, and J. Senthil, “IoT-based real time air pollution monitoring system,” *International Journal of Grid and High Performance Computing (IJGHPC)*, vol. 11, no. 4, pp. 28–41, 2019.
- [34] E. D. Ayele, K. Das, N. Meratnia, and P. J. Havinga, “Leveraging BLE and LoRa in IoT network for wildlife monitoring system (WMS),” in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, IEEE, 2018, pp. 342–348.
- [35] A. S. Syed, D. Sierra-Sosa, A. Kumar, and A. Elmaghraby, “IoT in smart cities: A survey of technologies, practices and challenges,” *Smart Cities*, vol. 4, no. 2, pp. 429–475, 2021.
- [36] A. K. Sikder, A. Acar, H. Aksu, A. S. Uluagac, K. Akkaya, and M. Conti, “IoT-enabled smart lighting systems for smart cities,” in *2018 IEEE 8th Annual Computing and Communication*

- Workshop and Conference (CCWC), IEEE, 2018, pp. 639–645.
- [37] A. Medvedev, P. Fedchenkov, A. Zaslavsky, T. Anagnostopoulos, and S. Khoruzhnikov, “Waste management as an IoT-enabled service in smart cities,” in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems: 15th International Conference, NEW2AN 2015, and 8th Conference, ruSMART 2015, St. Petersburg, Russia, August 26-28, 2015, Proceedings 15*, Springer, 2015, pp. 104–115.
- [38] S. A. Elsagheer Mohamed and K. A. AlShalfan, “Intelligent traffic management system based on the internet of vehicles (IoV),” *Journal of advanced transportation*, vol. 2021, pp. 1–23, 2021.
- [39] L.-M. Ang, K. P. Seng, G. K. Ijamaru, and A. M. Zungeru, “Deployment of IoV for smart cities: Applications, architecture, and challenges,” *IEEE access*, vol. 7, pp. 6473–6492, 2018.
- [40] M. Abdelsalam and T. Bonny, “IoV road safety: Vehicle speed limiting system,” in *2019 International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, IEEE, 2019, pp. 1–6.
- [41] Z. Yu, D. Jin, X. Song, C. Zhai, and D. Wang, “Internet of vehicle empowered mobile media scenarios: In-vehicle infotainment solutions for the mobility as a service (MaaS),” *Sustainability*, vol. 12, no. 18, p. 7448, 2020.
- [42] J. Li et al., “Battery-friendly relay selection scheme for prolonging the lifetimes of sensor nodes in the Internet of Things,” *IEEE Access*, vol. 7, pp. 33180–33201, 2019.
- [43] M. Radfar, A. Nakhlestani, H. Le Viet, and A. Desai, “Battery management technique to reduce standby energy consumption in ultra-low power IoT and sensory applications,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 1, pp. 336–345, 2019.
- [44] V. R. Konduru and M. R. Bharamagoudra, “Challenges and solutions of interoperability on IoT: How far have we come in resolving the IoT interoperability issues,” in *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, IEEE, 2017, pp. 572–576.
- [45] S. S. Albouq, A. A. Abi Sen, N. Almashf, M. Yamin, A. Alshanjiti, and N. M. Bahbouh, “A survey of interoperability challenges and solutions for dealing with them in IoT environment,” *IEEE Access*, vol. 10, pp. 36416–36428, 2022.
- [46] M. Noura, M. Atiquzzaman, and M. Gaedke, “Interoperability in internet of things infrastructure: Classification, challenges, and future work,” in *IoT as a Service: Third International Conference, IoTaaS 2017, Taichung, Taiwan, September 20–22, 2017, Proceedings 3*, Springer, 2018, pp. 11–18.
- [47] A. M. Yousuf, E. M. Rochester, B. Ousat, and M. Ghaderi, “Throughput, coverage and scalability of LoRa LPWAN for internet of things,” in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, IEEE, 2018, pp. 1–10.
- [48] J.-P. Bardyn, T. Melly, O. Seller, and N. Sorin, “IoT: The era of LPWAN is starting now,” in *ESSCIRC Conference 2016: 42nd European Solid-State Circuits Conference*, IEEE, 2016, pp. 25–30.
- [49] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, “A comparative study of LPWAN technologies for large-scale IoT deployment,” *ICT express*, vol. 5, no. 1, pp. 1–7, 2019.
- [50] X. Liu et al., “Transceiver design and multihop D2D for UAV IoT coverage in disasters,” *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1803–1815, 2018.
- [51] X. Li, H. Yao, J. Wang, S. Wu, C. Jiang, and Y. Qian, “Rechargeable multi-UAV aided seamless coverage for QoS-guaranteed IoT networks,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10902–10914, 2019.
- [52] Z. Song, K.-W. Chin, C. Yang, and M. Ros, “Methods to Assign UAVs for K-Coverage and Recharging in IoT Networks,” *IEEE Transactions on Mobile Computing*, 2023.
- [53] S. Khare and M. Totaro, “Big data in IoT,” in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, 2019, pp. 1–7.

- [54] Y. Sasaki, "A survey on IoT big data analytic systems: Current and future," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1024–1036, 2021.
- [55] S. P. Singh, A. Nayyar, R. Kumar, and A. Sharma, "Fog computing: from architecture to edge computing and big data processing," *The Journal of Supercomputing*, vol. 75, pp. 2070–2105, 2019.
- [56] T. Wang, H. Ke, X. Zheng, K. Wang, A. K. Sangaiah, and A. Liu, "Big data cleaning based on mobile edge computing in industrial sensor-cloud," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1321–1329, 2019.
- [57] M. Du, K. Wang, Y. Chen, X. Wang, and Y. Sun, "Big data privacy preserving in multi-access edge computing for heterogeneous Internet of Things," *IEEE Communications Magazine*, vol. 56, no. 8, pp. 62–67, 2018.
- [58] T. Mazhar et al., "Analysis of IoT security challenges and its solutions using artificial intelligence," *Brain Sciences*, vol. 13, no. 4, p. 683, 2023.
- [59] C. Alex, G. Creado, W. Almobaideen, O. A. Alghanam, and M. Saadeh, "A comprehensive survey for IoT security datasets taxonomy, classification and machine learning mechanisms," *Computers & Security*, p. 103283, 2023.
- [60] M. Poornima, H. Vimala, and J. Shreyas, "Holistic survey on energy aware routing techniques for IoT applications," *Journal of Network and Computer Applications*, vol. 213, p. 103584, 2023.
- [61] M. Z. Hussain and Z. M. Hanapi, "Efficient secure routing mechanisms for the low-powered IoT network: A literature review," *Electronics*, vol. 12, no. 3, p. 482, 2023.
- [62] R. Samadi, A. Nazari, and J. Seitz, "Intelligent energy-aware routing protocol in mobile IoT networks based on SDN," *IEEE Transactions on Green Communications and Networking*, 2023.
- [63] O. Aouedi and P. Kandaraj, "SURFS: Sustainable intrusion detection with hierarchical Federated Spiking neural networks," *ICC 2024*, 2024.
- [64] S.-U. Rehman, M. Khan, T. Zia, and L. Zheng, "Vehicular ad-hoc networks (VANETs): an overview and challenges," *Journal of Wireless Networking and communications*, vol. 3, no. 3, pp. 29–38, 2013.
- [65] A. M. Malla and R. K. Sahu, "A review on vehicle to vehicle communication protocols in VANETs," *International Journal*, vol. 3, no. 2, 2013.
- [66] M. Soni, B. S. Rajput, T. Patel, and N. Parmar, "Lightweight vehicle-to-infrastructure message verification method for VANET," in *Data Science and Intelligent Applications: Proceedings of ICDSIA 2020*, Springer, 2021, pp. 451–456.
- [67] H. Kathiriya, N. Kathiriya, and A. Bavarva, "Review on V2R Communication in VANET," in *Proceedings of the International Conference on Innovations in Automation and Mechatronics Engineering*, Vallabh Vidyanagar, India, 2013, pp. 21–23.
- [68] H. Ghaffarian, M. Soryani, and M. Fathy, "Planning VANET infrastructures to improve safety awareness in curved roads," *Journal of Zhejiang University SCIENCE C*, vol. 13, pp. 918–928, 2012.
- [69] X. Ma, J. Zhang, X. Yin, and K. S. Trivedi, "Design and analysis of a robust broadcast scheme for VANET safety-related services," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 46–61, 2011.
- [70] M. M. Hamdi, L. Audah, S. A. Rashid, and S. Alani, "VANET-based traffic monitoring and incident detection system: A review.," *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 11, no. 4, 2021.
- [71] S. B. Raut and L. Malik, "Survey on vehicle collision prediction in VANET," in *2014 IEEE International conference on computational intelligence and computing research*, IEEE, 2014, pp. 1–5.
- [72] Car production: Number of cars produced worldwide 2018 | Statista. <https://www.statista.com/statistics/262747/worldwide-automobile-production-since-2000/>,

Accessed 20 February 2024.

- [73] Oica (2020) World vehicles in use-all vehicles <https://www.oica.net/wp-content/uploads/Total-World-vehicles-in-use-2020.pdf>, Accessed 20 February 2024.
- [74] Richa, T. Sharma, and A. K. Sharma, "Heterogeneous-internet of vehicles (het-iov) in twenty-first century: a comprehensive study," *Handbook of computer networks and cyber security: principles and paradigms*, pp. 555–584, 2020.
- [75] L. Tuyisenge, M. Ayaida, S. Tohme, and L.-E. Afilal, "Network architectures in internet of vehicles (IoV): review, protocols analysis, challenges and issues," in *Internet of Vehicles. Technologies and Services Towards Smart City: 5th International Conference, IOV 2018, Paris, France, November 20–22, 2018, Proceedings 5*, Springer, 2018, pp. 3–13.
- [76] Bonomi, F.: The smart and connected vehicle and the Internet of Things. In: *Seminars, Cisco Systems: Advanced Architectures and Research, WSTS, San Jose* (2013). https://tf.nist.gov/seminars/WSTS/PDFs/1-0_Cisco_FBonomi_ConnectedVehicles.pdf, Accessed 20 March 2024.
- [77] R. Florin and S. Olariu, "A survey of vehicular communications for traffic signal optimization," *Vehicular Communications*, vol. 2, no. 2, pp. 70–79, 2015.
- [78] S. Sharma and B. Kaushik, "A survey on internet of vehicles: Applications, security issues & solutions," *Vehicular Communications*, vol. 20, p. 100182, 2019.
- [79] C.-C. Chang, Y.-M. Ooi, and B.-H. Sieh, "IoV-based collision avoidance architecture using machine learning prediction," *IEEE Access*, vol. 9, pp. 115497–115505, 2021.
- [80] A. Gruebler, K. D. McDonald-Maier, and K. M. A. Alheeti, "An intrusion detection system against black hole attacks on the communication network of self-driving cars," in *2015 sixth international conference on emerging security technologies (EST)*, IEEE, 2015, pp. 86–91.
- [81] R. Anil, M. Satyakumar, and A. Salim, "Emergency vehicle signal pre-emption system for heterogeneous traffic condition: A case study in trivandrum city," in *2019 4th International Conference on Intelligent Transportation Engineering (ICITE)*, IEEE, 2019, pp. 306–310.
- [82] Y. Sabri, A. Siham, and A. Maizate, "Internet of things (IoT) based smart vehicle security and safety system," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, 2021.
- [83] H. Mohapatra, A. K. Rath, and N. Panda, "IoT infrastructure for the accident avoidance: an approach of smart transportation," *International Journal of Information Technology*, vol. 14, no. 2, pp. 761–768, 2022.
- [84] L. Hu, J. Ou, J. Huang, Y. Chen, and D. Cao, "A review of research on traffic conflicts based on intelligent vehicles," *Ieee Access*, vol. 8, pp. 24471–24483, 2020.
- [85] C. Wang, S. Gong, A. Zhou, T. Li, and S. Peeta, "Cooperative adaptive cruise control for connected autonomous vehicles by factoring communication-related constraints," *Transportation Research Part C: Emerging Technologies*, vol. 113, pp. 124–145, 2020.
- [86] J. A. Onesimu, A. Kadam, K. M. Sagayam, and A. A. Elngar, "Internet of things based intelligent accident avoidance system for adverse weather and road conditions," *Journal of Reliable Intelligent Environments*, pp. 1–15, 2021.
- [87] R. Lu, X. Lin, H. Zhu, and X. Shen, "SPARK: A new VANET-based smart parking scheme for large parking lots," in *IEEE INFOCOM 2009*, IEEE, 2009, pp. 1413–1421.
- [88] F. Soldo, C. Casetti, C.-F. Chiasserini, and P. Chaparro, "Streaming media distribution in VANETs," in *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference*, IEEE, 2008, pp. 1–6.
- [89] A. Gholamhosseinian and J. Seitz, "A comprehensive survey on cooperative intersection management for heterogeneous connected vehicles," *IEEE Access*, vol. 10, pp. 7937–7972, 2022.
- [90] M. Roopa, S. A. Siddiq, R. Buyya, K. Venugopal, S. Iyengar, and L. Patnaik, "DTCMS: Dynamic

- traffic congestion management in Social Internet of Vehicles (SIoV),” *Internet of Things*, vol. 16, p. 100311, 2021.
- [91] J. Zhang and K. B. Letaief, “Mobile edge intelligence and computing for the internet of vehicles,” *Proceedings of the IEEE*, vol. 108, no. 2, pp. 246–261, 2019.
- [92] Y. Ni et al., “Toward reliable and scalable internet of vehicles: Performance analysis and resource management,” *Proceedings of the IEEE*, vol. 108, no. 2, pp. 324–340, 2019.
- [93] A. Arooj, M. S. Farooq, A. Akram, R. Iqbal, A. Sharma, and G. Dhiman, “Big data processing and analysis in internet of vehicles: architecture, taxonomy, and open research challenges,” *Archives of Computational Methods in Engineering*, vol. 29, no. 2, pp. 793–829, 2022.
- [94] X. Hou et al., “Reliable computation offloading for edge-computing-enabled software-defined IoV,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7097–7111, 2020.
- [95] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, “Connected vehicles: Solutions and challenges,” *IEEE internet of things journal*, vol. 1, no. 4, pp. 289–299, 2014.
- [96] Y. Ni, J. He, L. Cai, J. Pan, and Y. Bo, “Joint roadside unit deployment and service task assignment for Internet of Vehicles (IoV),” *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3271–3283, 2018.
- [97] J. Fan, L. K. Shar, J. Guo, W. Yang, D. Niyato, and K.-Y. Lam, “Differentiated security architecture for secure and efficient infotainment data communication in IoV networks,” in *International Conference on Network and System Security*, Springer, 2022, pp. 283–304.
- [98] R. Lu, X. Lin, H. Zhu, and X. Shen, “SPARK: A new VANET-based smart parking scheme for large parking lots,” in *IEEE INFOCOM 2009*, IEEE, 2009, pp. 1413–1421.
- [99] J. Ding, H. Mei, C.-L. I, H. Zhang, and W. Liu, “Frontier progress of unmanned aerial vehicles optical wireless technologies,” *Sensors*, vol. 20, no. 19, p. 5476, 2020.
- [100] X. Zhang and L. Duan, “Fast deployment of UAV networks for optimal wireless coverage,” *IEEE Transactions on Mobile Computing*, vol. 18, no. 3, pp. 588–601, 2018.
- [101] Project loon. <https://x.company/loon/>, Accessed 24 March 2024.
- [102] D. Gettinger, “Public safety drones,” Center for the study of drone at Bard College.(3rd ed.) Retrieved from: <https://dronecenter.bard.edu/files/2018/05/CSD-Public-Safety-Drones-Update-1.pdf>, 2020.
- [103] J. Zhang, W. Guo, B. Zhou, and G. S. Okin, “Drone-based remote sensing for research on wind erosion in drylands: possible applications,” *Remote Sensing*, vol. 13, no. 2, p. 283, 2021.
- [104] H. Hildmann and E. Kovacs, “Using unmanned aerial vehicles (UAVs) as mobile sensing platforms (MSPs) for disaster response, civil security and public safety,” *Drones*, vol. 3, no. 3, p. 59, 2019.
- [105] K. Rana, S. Praharaj, and T. Nanda, “Unmanned aerial vehicles (UAVs): An emerging technology for logistics,” *International Journal of Business and Management Invention*, vol. 5, no. 5, pp. 86–92, 2016.
- [106] S. Herwitz et al., “Imaging from an unmanned aerial vehicle: agricultural surveillance and decision support,” *Computers and electronics in agriculture*, vol. 44, no. 1, pp. 49–61, 2004.
- [107] M. T. Nguyen et al., “Uav-assisted data collection in wireless sensor networks: A comprehensive survey,” *Electronics*, vol. 10, no. 21, p. 2603, 2021.
- [108] L. Xie, X. Cao, J. Xu, and R. Zhang, “UAV-enabled wireless power transfer: A tutorial overview,” *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 4, pp. 2042–2064, 2021.
- [109] M. Abrar, U. Ajmal, Z. M. Almohaimed, X. Gui, R. Akram, and R. Masroor, “Energy efficient UAV-enabled mobile edge computing for IoT devices: A review,” *IEEE Access*, vol. 9, pp. 127779–127798, 2021.
- [110] S. A. H. Mohsan, M. A. Khan, M. H. Alsharif, P. Uthansakul, and A. A. A. Solyman, “Intelligent Reflecting Surfaces Assisted UAV Communications for Massive Networks: Current Trends, Challenges, and Research Directions,” *Sensors*, vol. 22, no. 14, 2022, doi: 10.3390/s22145278.

- [111] T. Akram, M. Awais, R. Naqvi, A. Ahmed, and M. Naeem, "Multicriteria UAV base stations placement for disaster management," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3475–3482, 2020.
- [112] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: Design, deployment and performance analysis," in *2015 IEEE global communications conference (GLOBECOM)*, IEEE, 2015, pp. 1–6.
- [113] I. Valiulahi and C. Masouros, "Multi-UAV deployment for throughput maximization in the presence of co-channel interference," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3605–3618, 2020.
- [114] S. Hosseinalipour, A. Rahmati, and H. Dai, "Interference avoidance position planning in UAV-assisted wireless communication," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, IEEE, 2019, pp. 1–6.
- [115] Y. Xue, B. Xu, W. Xia, J. Zhang, and H. Zhu, "Backhaul-aware resource allocation and optimum placement for UAV-assisted wireless communication network," *Electronics*, vol. 9, no. 9, p. 1397, 2020.
- [116] X. Sun and N. Ansari, "Jointly optimizing drone-mounted base station placement and user association in heterogeneous networks," in *2018 IEEE International Conference on Communications (ICC)*, IEEE, 2018, pp. 1–6.
- [117] E. Kalantari, M. Z. Shakir, H. Yanikomeroglu, and A. Yongacoglu, "Backhaul-aware robust 3D drone placement in 5G+ wireless networks," in *2017 IEEE international conference on communications workshops (ICC workshops)*, IEEE, 2017, pp. 109–114.
- [118] M. A. Ali and A. Jamalipour, "UAV placement and power allocation in uplink and downlink operations of cellular network," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4383–4393, 2020.
- [119] A. E. Duranay and H. Arslan, "The Evaluation of FFR for Interference Management in Coordinated Hybrid Terrestrial-Aerial Network," *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Istanbul, Turkey, 2019, pp. 1-6,
- [120] H. Pourrahmani, C. M. I. Bernier, and J. Van Herle, "The application of fuel-cell and battery technologies in unmanned aerial vehicles (UAVs): A dynamic study," *Batteries*, vol. 8, no. 7, p. 73, 2022.
- [121] C. Zhan and Y. Zeng, "Aerial-ground cost tradeoff for multi-UAV-enabled data collection in wireless sensor networks," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1937–1950, 2019.
- [122] S. Devasia and A. Lee, "Scalable low-cost unmanned-aerial-vehicle traffic network," *Journal of Air Transportation*, vol. 24, no. 3, pp. 74–83, 2016.
- [123] H. Zhao, H. Wang, W. Wu, and J. Wei, "Deployment algorithms for UAV airborne networks toward on-demand coverage," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 9, pp. 2015–2031, 2018.
- [124] Y. Zhi, Z. Fu, X. Sun, and J. Yu, "Security and privacy issues of UAV: A survey," *Mobile Networks and Applications*, vol. 25, no. 1, pp. 95–101, 2020.
- [125] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
- [126] W. Fang, W. Chen, W. Zhang, J. Pei, W. Gao, and G. Wang, "Digital signature scheme for information non-repudiation in blockchain: a state of the art review," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, pp. 1–15, 2020.
- [127] A. Kuznetsov, I. Oleshko, V. Tymchenko, K. Lisitsky, M. Rodinko, and A. Kolhatin, "Performance analysis of cryptographic hash functions suitable for use in blockchain," *International Journal of Computer Network & Information Security*, vol. 13, no. 2, pp. 1–15, 2021.
- [128] X. Fu, H. Wang, and P. Shi, "A survey of Blockchain consensus algorithms: mechanism, design and applications," *Science China Information Sciences*, vol. 64, pp. 1–15, 2021.

- [129] A. A. Monrat, O. Schelén, and K. Andersson, “A survey of blockchain from the perspectives of applications, challenges, and opportunities,” *Ieee Access*, vol. 7, pp. 117134–117151, 2019.
- [130] A. Reyna, C. Martín, J. Chen, E. Soler, and M. Díaz, “On blockchain and its integration with IoT. Challenges and opportunities,” *Future generation computer systems*, vol. 88, pp. 173–190, 2018.
- [131] A. Muneeza, N. A. Arshad, A. T. Arifin, and others, “The application of blockchain technology in crowdfunding: towards financial inclusion via technology,” *International journal of management and applied research*, vol. 5, no. 2, pp. 82–98, 2018.
- [132] A. Rahman et al., “Smartblock-sdn: An optimized blockchain-sdn framework for resource management in iot,” *IEEE Access*, vol. 9, pp. 28361–28376, 2021.
- [133] P. C. Franks, “Implications of blockchain distributed ledger technology for records management and information governance programs,” *Records Management Journal*, vol. 30, no. 3, pp. 287–299, 2020.
- [134] E. Hillbom and T. Tillström, “Applications of smart-contracts and smart-property utilizing blockchains,” 2016.
- [135] R. Yang et al., “Public and private blockchain in construction business process and information integration,” *Automation in construction*, vol. 118, p. 103276, 2020.
- [136] S. Bhasker, M. P. Grady, and K. G. Mosley, “Cryptocurrency and Anti-Money Laundering ENFORCEMENT AND REGULATION,” *Criminal Justice*, vol. 38, no. 2, pp. 3–11, 2023.
- [137] Q. Tao, X. Cui, X. Huang, A. M. Leigh, and H. Gu, “Food safety supervision system based on hierarchical multi-domain blockchain network,” *IEEE access*, vol. 7, pp. 51817–51826, 2019.
- [138] R. A. Michelin et al., “SpeedyChain: A framework for decoupling data from blockchain for smart cities,” in *Proceedings of the 15th EAI international conference on mobile and ubiquitous systems: Computing, networking and services*, 2018, pp. 145–154.
- [139] D. Morris, F. H. Sumner, and M. T. Wyld, “An appraisal of the Atlas Supervisor,” in *Proceedings of the 1967 22nd national conference*, 1967, pp. 67–75.
- [140] D. Doan Van and Q. Ai, “An efficient in-network caching decision algorithm for Internet of things,” *International Journal of Communication Systems*, vol. 31, no. 8, p. e3521, 2018.
- [141] J. Pfender, A. Valera, and W. K. Seah, “Easy as ABC: A lightweight centrality-based caching strategy for information-centric IoT,” in *Proceedings of the 6th ACM conference on information-centric networking*, 2019, pp. 100–111.
- [142] M. Amadeo, G. Ruggeri, C. Campolo, A. Molinaro, and G. Mangiullo, “Caching popular and fresh IoT contents at the edge via named data networking,” in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, IEEE, 2020, pp. 610–615.
- [143] H. Zhu, Y. Cao, X. Wei, W. Wang, T. Jiang, and S. Jin, “Caching transient data for Internet of Things: A deep reinforcement learning approach,” *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2074–2083, 2018.
- [144] M. Amadeo, C. Campolo, G. Ruggeri, and A. Molinaro, “Beyond edge caching: Freshness and popularity aware iot data caching via ndn at internet-scale,” *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 1, pp. 352–364, 2021.
- [145] X. Wu, X. Li, J. Li, P. Ching, V. C. Leung, and H. V. Poor, “Caching transient content for IoT sensing: Multi-agent soft actor-critic,” *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 5886–5901, 2021.
- [146] S. Wang, H. Chen, and Y. Wang, “Collaborative caching for energy optimization in content-centric Internet of Things,” *IEEE Transactions on Computational Social Systems*, vol. 9, no. 1, pp. 230–238, 2021.
- [147] M. I. A. Zahed, I. Ahmad, D. Habibi, and Q. V. Phung, “Content caching in industrial IoT: Security and energy considerations,” *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 491–504, 2019.
- [148] Z. Zhang, C.-H. Lung, I. Lambadaris, and M. St-Hilaire, “IoT data lifetime-based cooperative

- caching scheme for ICN-IoT networks,” in 2018 IEEE International Conference on Communications (ICC), IEEE, 2018, pp. 1–7.
- [149] B. Chen, L. Liu, M. Sun, and H. Ma, “IoTCache: Toward data-driven network caching for Internet of Things,” *IEEE Internet of Things journal*, vol. 6, no. 6, pp. 10064–10076, 2019.
- [150] D. Gupta, S. Rani, S. H. Ahmed, S. Verma, M. F. Ijaz, and J. Shafi, “Edge caching based on collaborative filtering for heterogeneous ICN-IoT applications,” *Sensors*, vol. 21, no. 16, p. 5491, 2021.
- [151] T.-V. Nguyen, N.-N. Dao, W. Noh, S. Cho, and others, “User-aware and flexible proactive caching using LSTM and ensemble learning in IoT-MEC networks,” *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 3251–3269, 2021.
- [152] Y. Hua, L. Guan, and K. G. Kyriakopoulos, “A fog caching scheme enabled by ICN for IoT environments,” *Future Generation Computer Systems*, vol. 111, pp. 82–95, 2020.
- [153] K. Hasan and S.-H. Jeong, “Efficient caching for data-driven IoT applications and fast content delivery with low latency in ICN,” *Applied Sciences*, vol. 9, no. 22, p. 4730, 2019.
- [154] J. Yao and N. Ansari, “Caching in dynamic IoT networks by deep reinforcement learning,” *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3268–3275, 2020.
- [155] R. Wang, Z. Kan, Y. Cui, D. Wu, and Y. Zhen, “Cooperative Caching Strategy With Content Request Prediction in Internet of Vehicles,” *IEEE Internet Things J.*, vol. 8, no. 11, pp. 8964–8975, Jun. 2021.
- [156] Z. Su, Y. Hui, Q. Xu, T. Yang, J. Liu, and Y. Jia, “An Edge Caching Scheme to Distribute Content in Vehicular Networks,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5346–5356, Jun. 2018.
- [157] C. Chen, J. Jiang, R. Fu, L. Chen, C. Li, and S. Wan, “An Intelligent Caching Strategy Considering Time-Space Characteristics in Vehicular Named Data Networks,” *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 10, pp. 19655–19667, Oct. 2022.
- [158] D. Gupta, S. Rani, S. H. Ahmed, S. Garg, Md. Jalil Piran, and M. Alrashoud, “ICN-Based Enhanced Cooperative Caching for Multimedia Streaming in Resource Constrained Vehicular Environment,” *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 7, pp. 4588–4600, Jul. 2021.
- [159] K. Zhang, J. Cao, S. Maharjan, and Y. Zhang, “Digital Twin Empowered Content Caching in Social-Aware Vehicular Edge Networks,” *IEEE Trans. Comput. Soc. Syst.*, vol. 9, no. 1, pp. 239–251, Feb. 2022.
- [160] C. Wang, C. Chen, Q. Pei, Z. Jiang, and S. Xu, “An Information-Centric In-Network Caching Scheme for 5G-Enabled Internet of Connected Vehicles,” *IEEE Trans. on Mobile Comput.*, vol. 22, no. 6, pp. 3137–3150, Jun. 2023.
- [161] C. Song, W. Xu, T. Wu, S. Yu, P. Zeng, and N. Zhang, “QoE-Driven Edge Caching in Vehicle Networks Based on Deep Reinforcement Learning,” *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5286–5295, Jun. 2021.
- [162] M. A. Javed and S. Zeadally, “AI-Empowered Content Caching in Vehicular Edge Computing: Opportunities and Challenges,” *IEEE Network*, vol. 35, no. 3, pp. 109–115, May 2021.
- [163] S. S. Musa, M. Zennaro, M. Libsle, and E. Pietrosemoli, “Mobility-Aware Proactive Edge Caching Optimization Scheme in Information-Centric IoV Networks,” *Sensors*, vol. 22, no. 4, p. 1387, Feb. 2022.
- [164] T. Z. H. Ernest and A. S. Madhukumar, “Ensemble Learning-Based Edge Caching Strategies for Internet of Vehicles: Outage and Finite SNR Analysis,” *IEEE Open J. Commun. Soc.*, vol. 4, pp. 239–252, 2023.
- [165] X. Zhou et al., “Edge Computation Offloading With Content Caching in 6G-Enabled IoV,” *IEEE Trans. Intell. Transport. Syst.*, pp. 1–15, 2024.
- [166] P. He, L. Cao, Y. Cui, R. Wang, and D. Wu, “Multi-Agent Caching Strategy for Spatial-Temporal Popularity in IoV,” *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13536–13546, Oct. 2023.

- [167] Z. Yu, J. Hu, G. Min, Z. Zhao, W. Miao, and M. S. Hossain, "Mobility-Aware Proactive Edge Caching for Connected Vehicles Using Federated Learning," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 8, pp. 5341–5351, Aug. 2021.
- [168] D. Zhang, W. Wang, J. Zhang, T. Zhang, J. Du, and C. Yang, "Novel Edge Caching Approach Based on Multi-Agent Deep Reinforcement Learning for Internet of Vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 24, no. 8, pp. 8324–8338, Aug. 2023.
- [169] M. Rahim, M. A. Javed, A. N. Alvi, and M. Imran, "An efficient caching policy for content retrieval in autonomous connected vehicles," *Transportation Research Part A: Policy and Practice*, vol. 140, pp. 142–152, Oct. 2020.
- [170] V. Sharma, I. You, D. N. K. Jayakody, D. G. Reina, and K.-K. R. Choo, "Neural-Blockchain-Based Ultrareliable Caching for Edge-Enabled UAV Networks," *IEEE Trans. Ind. Inf.*, vol. 15, no. 10, pp. 5723–5736, Oct. 2019.
- [171] A. Al-Hilo, M. Samir, C. Assi, S. Sharafeddine, and D. Ebrahimi, "UAV-Assisted Content Delivery in Intelligent Transportation Systems-Joint Trajectory Planning and Cache Management," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 8, pp. 5155–5167, Aug. 2021.
- [172] D. C. Nguyen et al., "Federated Learning Meets Blockchain in Edge Computing: Opportunities and Challenges," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12806–12825, Aug. 2021.
- [173] Y. Yin, M. Liu, G. Gui, H. Gacanin, H. Sari, and F. Adachi, "Cross-Layer Resource Allocation for UAV-Assisted Wireless Caching Networks With NOMA," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3428–3438, Apr. 2021.
- [174] T. Zhang, Z. Wang, Y. Liu, W. Xu, and A. Nallanathan, "Caching Placement and Resource Allocation for Cache-Enabling UAV NOMA Networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12897–12911, Nov. 2020.
- [175] S. Anokye, D. Ayepah-Mensah, A. M. Seid, G. O. Boateng, and G. Sun, "Deep Reinforcement Learning-Based Mobility-Aware UAV Content Caching and Placement in Mobile Edge Networks," *IEEE Systems Journal*, vol. 16, no. 1, pp. 275–286, Mar. 2022.
- [176] M. Zhang, M. El-Hajjar, and S. X. Ng, "Intelligent Caching in UAV-Aided Networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 739–752, Jan. 2022.
- [177] T. Zhang, Y. Wang, W. Yi, Y. Liu, and A. Nallanathan, "Joint Optimization of Caching Placement and Trajectory for UAV-D2D Networks," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5514–5527, Aug. 2022.
- [178] H. Wu, F. Lyu, C. Zhou, J. Chen, L. Wang, and X. Shen, "Optimal UAV Caching and Trajectory in Aerial-Assisted Vehicular Networks: A Learning-Based Approach," *IEEE J. Select. Areas Commun.*, vol. 38, no. 12, pp. 2783–2797, Dec. 2020.
- [179] J. Ji, K. Zhu, D. Niyato, and R. Wang, "Probabilistic Cache Placement in UAV-Assisted Networks With D2D Connections: Performance Analysis and Trajectory Optimization," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6331–6345, Oct. 2020.
- [180] Z. Su, M. Dai, Q. Xu, R. Li, and H. Zhang, "UAV Enabled Content Distribution for Internet of Connected Vehicles in 5G Heterogeneous Networks," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 8, pp. 5091–5102, Aug. 2021.
- [181] W. Wang, X. Xu, M. Bilal, M. Khan, and Y. Xing, "UAV-Assisted Content Caching for Human-Centric Consumer Applications in IoV," *IEEE Trans. Consumer Electron.*, pp. 1–1, 2024.
- [182] J. Luo, J. Song, F.-C. Zheng, L. Gao, and T. Wang, "User-Centric UAV Deployment and Content Placement in Cache-Enabled Multi-UAV Networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5656–5660, May 2022.
- [183] Y. Wang, C. Feng, T. Zhang, Y. Liu, and A. Nallanathan, "QoE Based Network Deployment and Caching Placement for Cache-Enabling UAV Networks," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, Dublin, Ireland: IEEE, Jun. 2020, pp. 1–6.

- [184] Y. Wang, S. Fu, C. Yao, H. Zhang, and F. R. Yu, "Caching Placement Optimization in UAV-Assisted Cellular Networks: A Deep Reinforcement Learning-Based Framework," *IEEE Wireless Commun. Lett.*, vol. 12, no. 8, pp. 1359–1363, Aug. 2023.
- [185] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, 2009, pp. 1–12.
- [186] S. Doi, S. Ata, H. Kitamura, and M. Murata, "IPv6 anycast for simple and effective service-oriented communications," *IEEE Communications Magazine*, vol. 42, no. 5, pp. 163–171, 2004.
- [187] M. Amadeo, C. Campolo, A. Molinaro, and G. Ruggeri, "Content-centric wireless networking: A survey," *Computer Networks*, vol. 72, pp. 1–13, 2014.
- [188] F. H. Lotfi and R. Fallahnejad, "Imprecise Shannon's entropy and multi attribute decision making," *Entropy*, vol. 12, no. 1, pp. 53–62, 2010.
- [189] B. R. Chandavarkar and R. M. R. Guddeti, "Simplified and improved analytical hierarchy process aid for selecting candidate network in an overlay heterogeneous networks," *Wireless Personal Communications*, vol. 83, pp. 2593–2606, 2015.
- [190] C. Bernardini, T. Silverston, and O. Festor, "Socially-aware caching strategy for content centric networking," in *2014 IFIP Networking Conference*, IEEE, 2014, pp. 1–9.
- [191] H.-T. Wu and G.-J. Horng, "Establishing an intelligent transportation system with a network security mechanism in an Internet of vehicle environment," *Ieee Access*, vol. 5, pp. 19239–19247, 2017.
- [192] R. Horvath, D. Nedbal, and M. Stieninger, "A literature review on challenges and effects of software defined networking," *Procedia Computer Science*, vol. 64, pp. 552–561, 2015.
- [193] B. Greenshields, J. Bibbins, W. Channing, and H. Miller, "A study of traffic capacity. In Highway research board proceedings, volume 1935," in *Highway research board proceedings*,
- [194] Mystakidis, Stylianos. "Metaverse." *Encyclopedia* 2, no. 1 (2022): 486-497.
- [195] Yang, Qiang, Yang Liu, Yong Cheng, Yan Kang, Tianjian Chen, and Han Yu. "Federated learning." *Synthesis Lectures on Artificial Intelligence and Machine Learning* 13, no. 3 (2019): 1-207.
- [196] W. Y. B. Lim et al., "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 2031–2063, 2020.
- [197] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings*, 2010, pp. 249–256.
- [198] R. DiPietro and G. D. Hager, "Deep learning: RNNs and LSTM," in *Handbook of medical image computing and computer assisted intervention*, Elsevier, 2020, pp. 503–519.
- [199] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in statistics: Methodology and distribution*, Springer, 1992, pp. 492–518.
- [200] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *Acm transactions on interactive intelligent systems (tiis)*, vol. 5, no. 4, pp. 1–19, 2015.
- [201] Z. Su, Y. Hui, and T. H. Luan, "Distributed task allocation to enable collaborative autonomous driving with network softwarization," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 10, pp. 2175–2189, 2018.
- [202] Z. Ding, P. Fan, G. K. Karagiannidis, R. Schober, and H. V. Poor, "NOMA assisted wireless caching: Strategies and performance analysis," *IEEE Transactions on Communications*, vol. 66, no. 10, pp. 4854–4876, 2018.
- [203] T. D. Wemegah, S. Zhu, and C. Atombo, "Modeling the effect of days and road type on peak period travels using structural equation modeling and big data from radio frequency identification for private cars and taxis," *European Transport Research Review*, vol. 10, pp. 1–14, 2018.

- [204] H. V. Abeywickrama, B. A. Jayawickrama, Y. He, and E. Dutkiewicz, “Comprehensive energy consumption model for unmanned aerial vehicles, based on empirical studies of battery performance,” IEEE access, vol. 6, pp. 58383–58394, 2018

