

Enhancing Smart City IoT Communication: A Two-Layer NOMA-Based Network with Caching Mechanisms and Optimized Resource Allocation

Tareq M. Alkhaldi, Abdulbasit A. Darem, Asma A. Alhashmi, Tawfik Al-Hadhrani, Azza Elneil Osman

Abstract—With advancements in next-generation communication systems, large-scale Internet of Things device (IoTds) deployments in smart cities face challenges like limited bandwidth, high latency, and network congestion. To address this, we propose a two-layer network architecture utilizing non-orthogonal multiple access (NOMA) and caching to enhance IoT communications' performance, efficiency, and reliability. Our primary objective is to optimize resource allocation and solve the association problem in a two-layer network. We formulated a joint optimization problem to maximize system utility through device association, power, and bandwidth allocation, considering constraints like channel quality and interference. We decoupled the non-linear, non-convex problem using block coordinate descent (BCD) and inner approximation techniques to maximize the aggregated data rate in high-density IoT scenarios. This approach reduced computational complexity while proving the scheme's theoretical and numerical convergence. To evaluate the proposed scheme, we compared its performance with an ideal backhauling approach, an exhaustive search (upper bound), and a Genetic algorithm-based heuristic. Our scheme outperformed the others, achieving 98.04% of the ideal backhauling and 99.60% of the upper bound. Statistical analysis confirmed its robustness and consistent performance across various conditions. The two-layer NOMA-based network with caching and optimized resource allocation significantly enhances IoT communication efficiency and resilience, offering a solid framework for future smart city deployments.

Index Terms—Non-Orthogonal Multiple Access (NOMA) Heterogeneous networks (HetNets) Resource allocation Spectral efficiency Network latency

I. INTRODUCTION

Recent advancements in next-generation 5G and beyond 5G wireless communication systems provide benefits such as ex-

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2024- 2913-09". This study is supported via funding from Prince sattam bin Abdulaziz University project number (PSAU/2024/R/1445). This research project was conducted jointly with the Computer Science Department at Nottingham Trent University, United Kingdom.

Tareq M. Alkhaldi is with the Department of Educational Technologies, Imam Abdulrahman bin Faisal University, Dammam 4221, Saudi Arabia. Email: talkhaleidi@iau.edu.sa

Abdulbasit A. Darem and Asma A. Alhashmi are with the Department of Computer Science at college of Science, Northern Border University, Arar, Saudi Arabia. Email: Basit.darem@nbu.edu.sa & asma.alhashmi@nbu.edu.sa

Tawfik Al-Hadhrani is in the Computer Science Department, School of Science and Technology, Nottingham Trent University, NG11 8NS, UK. Tawfik.al-hadhrani@ntu.ac.uk

Azza Elneil Osman is with the Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia. a.abdalla@psau.edu.sa

Corresponding author: Asma A. Alhashmi, asma.alhashmi@nbu.edu.sa

tensive connectivity, ultra-reliable low-latency communication, and high data rates [1], [2]. However, the extensive deployment of IoT devices in smart cities poses significant challenges for traditional networks due to limited power resources and finite bandwidth, leading to high latency and network congestion, which hinder the efficient deployment of IoT applications [3]. An essential component of modern IoT applications, like smart traffic management, is reliable backhaul communication, ensuring efficient real-time data transmission from numerous sensors to the central controller (CC) [4]. Traditional backhaul networks, constrained by limited computational and communication resources, result in high latency and increased susceptibility to interference, obstructing efficient communication [5]. To address the limitations of traditional networks, an effective multilayer communication network is needed, where the first layer connects the IoT devices to the base stations (BS), and the second layer links the BS and IoT to a CC [6]. Through the use of this multilayer architecture, the traditional network can offer the low latency and high capacity of terrestrial networks, coupled with caching at base stations (BSs) to store frequently accessed data closer to users [7]. Consequently, a multilayer network, combined with an integrated caching approach at the BSs, enhances overall network performance, providing a more efficient, reliable, and scalable solution critical for the successful implementation of smart cities [8]. Additionally, this network requires significant interference management strategies between devices and between different layers of the communication network [9], [10].

In addition to the multilayer architecture, the choice of multiple access techniques is crucial. Traditional approaches, such as orthogonal multiple access (OMA), have been prominent for many years [11]. However, OMA is not well-suited for applications constrained by limited spectrum [12], leading to under-utilized spectrum and reduced network capacity, particularly problematic in densely populated IoT environments [8]. Nonorthogonal multiple access (NOMA) offers a promising alternative [13], allowing multiple users to share the same frequency resources simultaneously by using different power levels, thereby enhancing spectral efficiency [14]. Furthermore, NOMA improves fairness among IoT devices by dynamically adjusting power based on channel conditions, making it highly suitable for real-time IoT applications [15]. By integrating NOMA into a multilayer network, the overall performance of smart cities is significantly improved in terms of reliability, efficiency, and the effectiveness of IoT devices [16]. Additionally, incorporating a caching system at BSs in combination with NOMA enhances data transmission by storing frequently

requested information [7], [17]. In the following section, we will explore recent advancements in these technologies both individually and in an integrated manner.

II. RECENT ADVANCEMENTS

To overcome the challenges mentioned above, several works related to smart city applications have been published in the literature and address the significant challenges raised due to the dense deployment of IoT devices, as well as the limited power and bandwidth that result in high latency and network congestion. Consequently, the work done by Al-Abbasi et al. [18], Zhang et al. [19], and Vaezi et al. [1], [20], mention the shortcomings related to finite computational and communication resources of dense network deployed in the smart cities. Following that, the work done by Lee et al. [13] and Lim et al. [21] emphasizes the need for new and promising solutions for IoT data management that require reliable backhaul networks. These networks facilitate data transmission and reception between BSs and the core network. In addition to these benefits, the traditional network encounters issues, for example, limited capacity, high latency, and vulnerability to cyberattacks, as demonstrated by the work of Kwon et al. [22], Alameer Ahmad et al. [23] and Kaneko et al. [14]. Therefore, to overcome these challenges, a considerable amount of effort is required to pave the way toward a multilayer communication network. This network links IoT devices with BS in the first layer and BSs with CC in the second layer, resulting in operational efficiency. Studies by Gevez et al. [12], Wu et al. [24], and Chrysologou et al. [25] have demonstrated the success of these architectures. Furthermore, Yang et al. [17] and Hao et al. [26] have explored the integration of fog and cloud computing within hybrid architectures, which improves the network's ability to handle extensive IoT projects.

The literature has extensively studied traditional multiple access methods such as OMA for the IoT network, in addition to multilayer architecture. However, they show inherent inefficiencies in resource utilization, leading to an underutilized spectrum and reduced network capacity, especially in densely populated IoT environments. In contrast to the advantages of OMA, it also faces some limitations in handling large data traffic in dense urban environments such as smart cities, as discussed by Al-Abbasi et al. [18], Zhang et al. [19] and Lee et al. [9]. To overcome the limitation, NOMA technology presents a promising alternative that allows multiple users to share the same frequency resources simultaneously using different power levels, enhancing spectral efficiency. Therefore, with efficient utilization of NOMA, network performance improves significantly, resulting in reduced latency and improved user fairness, making it a good candidate for real-time IoT applications, as explained in research by Won et al. [22], Lee et al. [9], and Moltafet et al. [27]. Moreover, to meet the requirements of massive connectivity and high spectral efficiency for real-time applications, NOMA became the best candidate, according to Zhang et al. [28] and Bohli et al. [29]. Moreover, recent advances in NOMA, as highlighted by Gevez et al. [12] and Lim et al. [30], pave the way for its integration with machine learning techniques for adaptive resource management.

Given NOMA's benefits, integrating it into a multi-layer network can improve IoT device performance, including spectral efficiency and reliability, in urban smart environments [31]. Besides, BS caching strategies also enhance data transmission by storing frequently accessed data, reducing redundant transmissions and network congestion [30]–[32]. Therefore, traditional communication problems, as discussed earlier, like the need for effective bandwidth and power limitations, can be overcome by integrating the NOMA with caching in the future communication network, as confirmed by the studies done by Yang et al. [17], Hao et al. [26], and Guo et al. [7]. Moreover, the studies done by the authors of [22] and [29] further enhance the effectiveness of NOMA-enabled caching in multilayer networks and address the requirements of future communication systems. Besides these advances, most of them work mainly on either the fronthaul or the backhaul, leaving the combined effect unexplored. To fill this gap, our work investigates NOMA-enabled caching under quality of service constraints and dynamic resource allocation in dual-layer networks. Besides this, most of the work overlooks the potential benefits of integrating these technologies for more resilient and scalable networks, particularly in smart cities. Dynamic resource allocation, considering real-time data demands and fluctuating network conditions, remains a challenge. To tackle IoT challenges in smart cities, a comprehensive strategy must include NOMA and caching, optimized within a dual-layer network framework, which is the main novelty of this work. As per my best knowledge, this is the first work done and lays the groundwork for future communication networks.

A. Motivation and Contributions

Building on the previous discussion, this work considers NOMA along with caching at the BSs by introducing an innovative multilayer network architecture. The first layer links IoT devices to multiple BSs, while the second layer connects BSs/IoTs to a central controller. The proposed architecture overcomes single-layer architecture limitations by integrating the NOMA in the network to improve spectral efficiency and reduce cross-tier interference induced by the multilayer network. Following that, caching at the BSs further reduces redundant data transmissions and network congestion, improving the system. This integrated approach meets the growing bandwidth and data rate needs of IoT applications, ensuring seamless and efficient communication in smart cities. Moreover, efficient and dynamic allocation of communication resources will further minimize interference and increase data rate as one of the main contributions of this work for multilayer networks. Thus, the proposed framework results in a scalable and efficient network architecture that can meet smart city data demands, improve existing solutions, and fill critical gaps as discussed above. In addition, the proposed model aims to pave the way for more resilient and efficient IoT networks, ultimately improving the quality of service and user experience in smart cities. The main contributions of this work are summarized as follows:

- 1) This work formulates the joint optimization problem for resource allocation in a multi-layer network that

integrates NOMA and caching mechanisms, addressing the unique challenges of smart city IoT environments while taking into account the cross-tier interference. The optimization problem is designed to maximize the aggregated data rate, ensuring efficient resource utilization in high-density IoT scenarios by jointly optimizing the user associations and communication resources.

- 2) To address the complexity of the joint optimization problem, we decoupled the joint optimization problem into subproblems using the block coordinate (BCD) methods and transformed it into a more manageable form using the inner approximation methods, resulting in a reduction in overall computational complexity. Furthermore, to demonstrate the effectiveness of the proposed scheme, simulations were conducted using an exhaustive search which poses the upper bound for our solution and also with a heuristic approach based on the genetic algorithm.
- 3) The incorporation of caching at base stations in the proposed architecture effectively reduces redundant data transmissions and alleviates network congestion, enhancing overall network efficiency. The results demonstrate that the proposed scheme outperforms the others, while performing 98.04% equal to the ideal backhauling and approximately 99.60% equal to the upper bound solution.
- 4) Following that to further reveal the effectiveness of the proposed scheme and further support the numerical findings, the Statistical analysis was also conducted along with the simulations to, highlighting the consistent performance of the proposed scheme under varying network conditions and its robustness compared to other methods. This analysis shows that the proposed scheme maintains superior performance across different scenarios, reinforcing its suitability for practical applications.

To improve clarity and ease of reference, Table I lists the key notations used throughout the paper.

III. SYSTEM MODEL AND MATHEMATICAL MODEL

This section introduces a two-layer caching-based system that leverages NOMA technology to enhance IoT communication within a smart city. Our focus lies primarily on the uplink communication scenario within this framework, characterized by a network comprising M base stations (BS) and a central controller (CC) denoted as k , which serves Q IoT devices (IoT-D) as shown in Fig. 1.¹ BSs are crucial for seamless communication between IoT devices and the CC in the smart city. Without them, IoT devices would rely solely on their own communication, risking breakdowns and compromising network efficiency. For effective communication, the system bandwidth, represented as β , is divided into two segments: front-haul and backhaul links. Here, $(1 - \lambda)B$ is allocated to front-haul communication between the IoT-Ds and the BSs, while λB is dedicated to backhaul communication between the

TABLE I: List of Notations

Symbol	Description
M	Number of base stations (BSs)
k	Central controller (CC)
Q	Number of IoT devices (IoT-Ds)
β	Total system bandwidth
λ	Bandwidth allocation coefficient
$P_{q,m}$	Transmission power of IoT-D q to BS m
$P_{q,k}$	Transmission power of IoT-D q to CC
$H_{q,m}$	Channel gain between IoT-D q and BS m
$H_{q,k}$	Channel gain between IoT-D q and CC
ζ_q	Additive white Gaussian noise (AWGN)
σ^2	Noise variance
$R_{q,m}$	Data rate from IoT-D q to BS m
$R_{q,k}$	Data rate from IoT-D q to CC
$L_{q,m}$	Large-scale path loss between IoT-D q and BS m
$L_{q,k}$	Large-scale path loss between IoT-D q and CC
α	Path loss exponent
$d_{q,m}$	Distance between IoT-D q and BS m
$d_{q,k}$	Distance between IoT-D q and CC
d_0	Reference distance
$\gamma_{q,m}$	SINR for IoT-D q and BS m
$\gamma_{q,k}$	SINR for IoT-D q and CC
Φ_m^{inter}	Inter-cell interference at BS m
Φ_m^{intra}	Intra-cell interference at BS m
I_m	Total interference at BS m
$\nu_{q,m}$	Logarithmic term for SINR of IoT-D q and BS m
η	Cache hit ratio
$\mathbf{A}$	Device association matrix
$\mathbf{P}$	Power allocation matrix
B	Total bandwidth
R_m	Overall achievable rate at BS m
$R_m^{\text{cache,SIC}}$	Achievable rate at BS m with caching and SIC

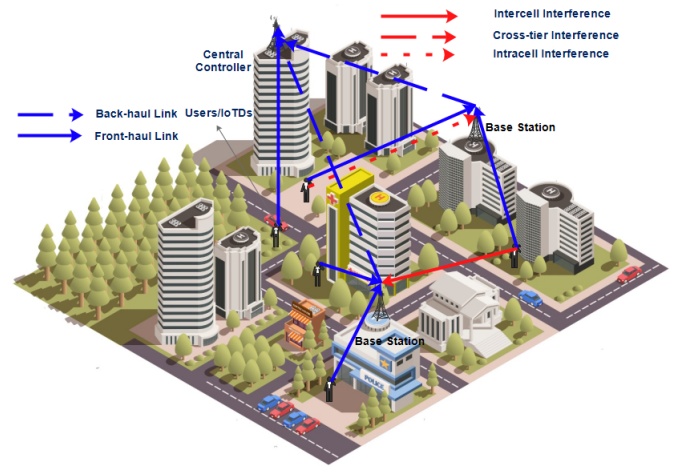


Fig. 1: The proposed two-layer caching-based network model for IoT communication in smart cities using NOMA technology. IoT-Ds connect to BSs, which link to a CC. The model optimizes uplink communication through bandwidth allocation, caching, and interference management.

BSs and the CC. To effectively address network requirements, the system incorporates wireless caching technology within BSs, enabling data offloading. Furthermore, NOMA technology facilitates multi-device transmission, enhancing system efficiency. Communication between BSs and the CC primarily occurs over the C-band. Moreover, the system model includes different types of cochannel interference, such as cross-tier CC interference, intercell interference, and intracell interference from NOMA. Subsequent sections provide a detailed analysis

¹While our focus is on optimizing communication, we assume the network is fully secured [33]. Future work will consider integrating security mechanisms such as encryption and secure authentication protocols.

of the system's performance and design, highlighting the importance of our approach in optimizing network resources and improving overall performance.

A. Communication Model

Furthermore, the communication model plays a crucial role in demonstrating the effectiveness of the proposed scheme. Therefore, it needs to be addressed efficiently, as explained in the subsequent section involving terrestrial communication systems, which include the front haul from IoTs to BS and the backhaul from IoTs/BS to the CC.

1) *Terrestrial Communication Model*: In the context of HetNets enabled by Smart City IoT, NOMA technology allows a BS to serve multiple IoTs concurrently. Consider a set of IoTs, indexed by q , served by a particular BS m . The received signal at BS m from IoT q can be modeled as:

$$y_{qm} = \sqrt{P_{q,m}} H_{q,m} S_{q,m} + \sum_{q' \neq q} \sqrt{P_{q',m}} H_{q',m} S_{q',m} + I_m + \zeta_q, \quad (1)$$

In eqn. (1), $P_{q,m}$ denotes the transmission power of IoT q to BS m , $H_{q,m}$ represents the channel gain between IoT q and BS m , and $S_{q,m}$ is the transmitted signal. The interference term I_m encompasses both inter-cell and intra-cell interference components: $I_m = \Phi_m^{\text{inter}} + \Phi_m^{\text{intra}}$, with $\Phi_m^{\text{inter}} = \sum_{j \neq m} \sum_i \sqrt{P_{i,j}} H_{i,m} S_{i,j}$, $\Phi_m^{\text{intra}} = \sum_{q' \neq q} \sqrt{P_{q',m}} H_{q',m} S_{q',m}$. Here, ζ_q denotes the additive white Gaussian noise (AWGN) with $\zeta_q \sim \mathcal{CN}(0, \sigma^2)$, where σ^2 is the noise variance. The communication channel between IoT q and BS m is modeled as a Rayleigh fading channel: $H_{q,m} = \sqrt{L_{q,m}} h_{q,m}$, where $L_{q,m}$ represents the large-scale path loss and $h_{q,m} \sim \mathcal{CN}(0, 1)$ represents the small-scale fading coefficient. The large-scale path loss $L_{q,m}$ is typically modeled as a function of the distance $d_{q,m}$ between IoT q and BS m : $L_{q,m} = \left(\frac{d_0}{d_{q,m}}\right)^\alpha$, where d_0 is the reference distance, $d_{q,m}$ is the actual distance between IoT q and BS m , and α is the path loss exponent, which typically ranges between 2 (free space) and 4 (urban environments). The signal-to-interference-plus-noise ratio (SINR) for the communication link between IoT q and BS m is defined as:

$$\Gamma_{q,m} = \frac{P_{q,m} |H_{q,m}|^2}{\sum_{q' \neq q} P_{q',m} |H_{q',m}|^2 + I_m + \sigma^2}, \quad (2)$$

where $P_{q,m}$ is the transmission power, $|H_{q,m}|^2$ is the channel gain, and I_m represents the interference. The bandwidth allocation coefficient λ determines the division of the total bandwidth β between front-haul and backhaul links. Specifically, $(1 - \lambda)\beta$ is designated for front-haul communication, while $\lambda\beta$ is allocated for backhaul communication. The device association matrix $\mathbf{A} = [a_{q,m}]_{Q \times (M+1)}$ defines the association of IoTs with BS/CC. If IoT q is associated with BS/CC m , then $a_{q,m} = 1$; otherwise, $a_{q,m} = 0$. Following that, the transmission rate from IoT q to BS m is given by:

$$R_{q,m} = (1 - \lambda)\beta \log_2(1 + \gamma_{q,m}). \quad (3)$$

The overall achievable rate at BS m is the summation of the rates of all associated IoTs: $R_m = \sum_q a_{q,m} R_{q,m}$.

2) *Caching and SIC Decoding Mechanism*: To further improve the efficiency of our system, we integrate caching and SIC decoding mechanisms at the BS. The caching mechanism stores popular content at the BS, reducing redundant data transmissions and alleviating network congestion. The cache hit ratio determines the effectiveness of caching, indicated as η . The cache hit ratio represents the probability that a requested content is available in the cache.

The received signal at the BS m after SIC decoding is given by:

$$y_{qm}^{\text{SIC}} = \sqrt{P_{q,m}} H_{q,m} S_{q,m} + \sum_{q' \in \mathcal{U}_q} \sqrt{P_{q',m}} H_{q',m} S_{q',m} + I_m + \zeta_q, \quad (4)$$

where $\mathcal{U}_q$ represents the set of users whose signals are decoded before user q 's signal. Users' interference in $\mathcal{U}_q$ is successively canceled, improving the SINR for the user q .

The SINR after SIC for user q at BS m is given by:

$$\Gamma_{q,m}^{\text{SIC}} = \frac{P_{q,m} |H_{q,m}|^2}{\sum_{q' \notin \mathcal{U}_q} P_{q',m} |H_{q',m}|^2 + I_m + \sigma^2}, \quad (5)$$

The achievable rate after SIC for user q at BS m is:

$$R_{q,m}^{\text{SIC}} = (1 - \lambda)\beta \log_2(1 + \Gamma_{q,m}^{\text{SIC}}). \quad (6)$$

The overall achievable rate at BS m with caching and SIC is then given by:

$$R_m^{\text{cache,SIC}} = \eta \sum_{q \in \mathcal{Q}} a_{q,m} R_{q,m}^{\text{SIC}} + (1 - \eta) \sum_{q \in \mathcal{Q}} a_{q,m} R_{q,m}. \quad (7)$$

This model integrates the benefits of caching and SIC to enhance the overall system performance, reduce latency, and improve data rates for IoT devices in the smart city HetNet.

3) *IoT-CC Communication*: The received signal at the CC from the q^{th} IoT, denoted as $y_{q,k}$, is modeled as:

$$y_{q,k} = \sqrt{P_{q,k}} H_{q,k} S_{q,k} + I_{q,k} + \zeta_k, \quad (8)$$

where $P_{q,k}$ represents the transmission power of IoT q to the CC, $H_{q,k}$ is the channel gain between IoT q and the CC, $S_{q,k}$ is the transmitted signal, $I_{q,k}$ denotes the total interference, and $\zeta_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise with variance σ^2 . The total interference $I_{q,k}$ consists of intra-cell and inter-cell components, defined as: $I_{q,k} = \Phi_{q,k}^{\text{intra}} + \Phi_{q,k}^{\text{inter}}$, where: $\Phi_{q,k}^{\text{intra}} = \sum_{q' \neq q} \sqrt{P_{q',k}} H_{q',k} S_{q',k}$, $\Phi_{q,k}^{\text{inter}} = \sum_{m \neq k} \sum_i \sqrt{P_{i,m}} H_{i,k} S_{i,m}$. The channel gain $H_{q,k}$ encompasses both large-scale path loss and small-scale Rayleigh fading: $H_{q,k} = \sqrt{L_{q,k}} h_{q,k}$, where $L_{q,k}$ represents the large-scale path loss and $h_{q,k} \sim \mathcal{CN}(0, 1)$ denotes the small-scale fading coefficient. The large-scale path loss $L_{q,k}$ as a function of the distance $d_{q,k}$ between IoT q and the CC is given by: $L_{q,k} = \left(\frac{d_0}{d_{q,k}}\right)^\alpha$, where d_0 is the reference distance, $d_{q,k}$ is the actual distance, and α is the path loss exponent. The SINR $\Gamma_{q,k}$ for the link between IoT q and the CC is defined as:

$$\Gamma_{q,k} = \frac{P_{q,k} |H_{q,k}|^2}{\Phi_{q,k}^{\text{intra}} + \Phi_{q,k}^{\text{inter}} + \sigma^2}. \quad (9)$$

Finally, the achievable data rate $R_{q,k}$ at the CC from IoT q is determined by:

$$R_{q,k} = \lambda\beta \log_2(1 + \Gamma_{q,k}), \quad (10)$$

where λ is the bandwidth allocation coefficient and β is the total bandwidth.

B. Problem Formulation

Building on the previous discussion, this section aims to formulate the joint optimization problem to maximize the system's utility by optimizing device association $\mathbf{A}$, power allocation $\mathbf{P}$, and bandwidth allocation λ , while adhering to the constraints of the system and communication. These constraints are influenced by channel quality, interference, and available bandwidth. The joint optimization problem is formulated as follows:

$$\text{maximize}_{\mathbf{A}, \mathbf{P}, \lambda} \sum_{q \in \mathbb{Q}} \left(\sum_{m \in \mathbb{M}} a_{q,m} R_{q,m}^{\text{cache,SIC}} + a_{q,k} R_{q,k} \right), \quad (11a)$$

$$\text{subject to: } \sum_{q \in \mathbb{Q}} a_{q,m} P_{q,m} \leq P_m^{\max}, \quad \forall m \in \mathbb{M}, \quad (11b)$$

$$\sum_{q \in \mathbb{Q}} a_{q,k} P_{q,k} \leq P_k^{\max}, \quad (11c)$$

$$a_{q,m} \in \{0, 1\}, \quad \forall q \in \mathbb{Q}, \forall m \in \mathbb{M} \cup \{k\}, \quad (11d)$$

$$\sum_{m \in \mathbb{M} \cup \{k\}} a_{q,m} = 1, \quad \forall q \in \mathbb{Q}, \quad (11e)$$

$$0 \leq \lambda \leq 1, \quad (11f)$$

$$\gamma_{q,m}^{\text{SIC}} \geq \gamma_{q,m}^{\min}, \quad \forall q \in \mathbb{Q}, \forall m \in \mathbb{M}, \quad (11g)$$

$$R_{q,m}^{\text{SIC}} \geq R_{q,m}^{\min}, \quad \forall q \in \mathbb{Q}, \forall m \in \mathbb{M}, \quad (11h)$$

The device association matrix $\mathbf{A}$, power allocation matrix $\mathbf{P}$, and bandwidth allocation coefficient λ are defined. The objective function in eqn. (11a) maximizes the aggregate data rate by summing the rates of all IoTs to their respective BS or CC. The constraint in eqn. (11b) ensures that the total power allocated to each BS does not exceed $P_m^{\max}$, and constraint in eqn. (11c) imposes a similar limit for the CC. The binary constraint in eqn. (11d) guarantees that each IoT is associated with exactly one BS or CC and constraint in eqn. (11e) ensures that each IoT connects to only one BS or the CC. The bandwidth allocation constraint in eqn. (11f) restricts λ to $[0, 1]$, ensuring a valid distribution. To improve performance, constraints in eqn. (11g) ensures SINR $\gamma_{q,m}^{\text{SIC}}$ for each IoT q associated with BS m meets the minimum $\gamma_{q,m}^{\min}$, and constraint eqn. (11h) guarantees that the data rate $R_{q,m}^{\text{SIC}}$ for each IoT q associated with BS m exceeds $R_{q,m}^{\min}$. These constraints ensure optimal resource allocation while meeting system requirements and performance thresholds.

1) Remark on the Nature of the Optimization Problem:

The optimization problem formulated is inherently complex due to its mixed-integer nonlinear programming (MINLP) nature, involving both continuous variables (power and bandwidth allocation) and binary variables (device association). The nonlinearity stems from the SINR expressions, ratios of quadratic terms that complicate the solution process. The mixed integer aspect of the problem arises from the binary device association matrix $\mathbf{A}$ and the continuous variables in the power allocation matrix $\mathbf{P}$ and bandwidth allocation λ . This combination significantly increases complexity. Additionally, SINR and rate constraints are nonconvex due to interference

terms and logarithmic functions, making it difficult to find a global optimum. The problem is NP-hard, indicating that no known polynomial-time algorithm exists for an exact solution, mainly due to the combinatorial nature of device association and the non-convexity of the mathematical functions.

The SINR for the link between IoT q and BS m is $\gamma_{q,m} = \frac{P_{q,m} |H_{q,m}|^2}{\sum_{q' \neq q} P_{q',m} |H_{q',m}|^2 + I_m + \sigma^2}$, confirming the nonlinearity with its ratio of quadratic terms. The rate $R_{q,m} = (1-\lambda)B \log_2(1 + \gamma_{q,m})$ involves a logarithmic function of SINR, contributing to nonlinearity. The binary variables $a_{q,m} \in \{0, 1\}$ and continuous variables $P_{q,m}$ and λ classify the problem as mixed-integer, increasing its complexity. The nonconvex objective function $\sum_{q \in \mathbb{Q}} (\sum_{m \in \mathbb{M}} a_{q,m} R_{q,m} + a_{q,k} R_{q,k})$ involves maximizing the sum of logarithmic functions of SINRs, further complicating the problem. The NP-hard nature is due to the combinatorial complexity of device association and the nonconvex SINR and rate constraints. Therefore, to address this, in the following section, we decouple the original optimization problem in eqn (11) into a series of subproblems and solve them iteratively to find near-optimal solutions within reasonable computational times.

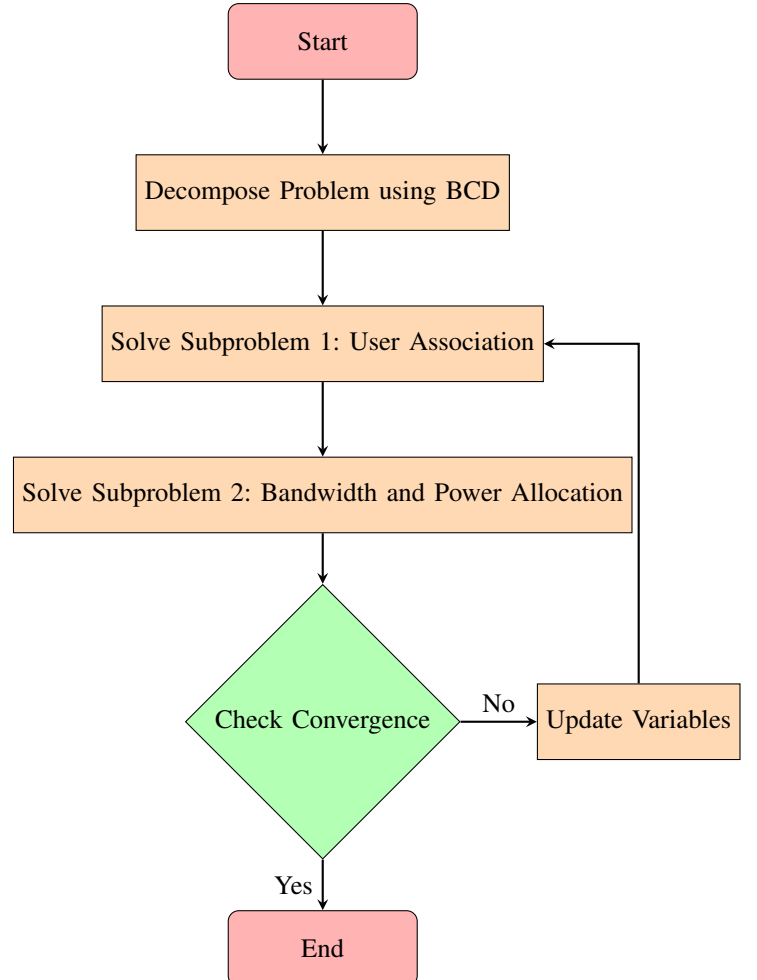


Fig. 2: Flowchart of the Decoupled Problem Solving Process

IV. PROPOSED FRAMEWORK

This section introduces the proposed framework for solving eqn. (11), considering the characteristics of the optimization problem as discussed in Section III-B1. To tackle this, we decompose the original optimization problem using the block coordinate descent (BCD) method into two subproblems: user association Section IV-A and bandwidth and power allocation Section IV-B. This decomposition not only simplifies the problem but also allows for more efficient and tractable solutions with low complexity, thereby enhancing the system's efficiency. Additionally, Fig.2 illustrates the flowchart that details the entire operational process.

A. Subproblem 1: User Association

Under the fixed bandwidth and power allocation values, the first subproblem focuses on determining the optimal association of IoTs with BSs/CC. This can be expressed as follows:

$$\underset{\mathbf{A}}{\text{maximize}} \sum_{q \in \mathbb{Q}} \left(\sum_{m \in \mathbb{M}} a_{q,m} R_{q,m}^{\text{cache,SIC}} + a_{q,k} R_{q,k} \right), \quad (12a)$$

$$\text{subject to: } a_{q,m} \in \{0, 1\}, \quad \forall q \in \mathbb{Q}, \forall m \in \mathbb{M} \cup \{k\}, \quad (12b)$$

$$\sum_{m \in \mathbb{M} \cup \{k\}} a_{q,m} = 1, \quad \forall q \in \mathbb{Q}. \quad (12c)$$

This subproblem aims to maximize the aggregate data rate by optimizing the device association matrix $\mathbf{A}$. Constraints in eqn. (12b) and in eqn. (12c) ensure that each IoT is associated with exactly one BS or the CC.

Note: The optimization problem in eqn. (12) is still non-convex and nonlinear. Thus, finding the optimal solution remains computationally challenging. Therefore, to address this, we transform it into a more tractable form by relaxing the binary variable to be continuous between 0 and 1.

By relaxing the binary variable $a_{q,m}$ to be continuous between 0 and 1, we obtain the following relaxed optimization problem:

$$\underset{\mathbf{A}}{\text{maximize}} \sum_{q \in \mathbb{Q}} \left(\sum_{m \in \mathbb{M}} a_{q,m} R_{q,m}^{\text{cache,SIC}} + a_{q,k} R_{q,k} \right), \quad (13a)$$

$$\text{subject to: } 0 \leq a_{q,m} \leq 1, \forall q \in \mathbb{Q}, \forall m \in \mathbb{M} \cup \{k\}, \quad (13b)$$

$$\sum_{m \in \mathbb{M} \cup \{k\}} a_{q,m} = 1, \quad \forall q \in \mathbb{Q}. \quad (13c)$$

This relaxation transforms the problem into a more tractable linear programming (LP) problem that can be efficiently solved using standard optimization techniques. After solving the relaxed problem, we can apply a rounding technique to convert the continuous variables to binary form, ensuring a feasible solution to the original problem. A detailed explanation and the working principle are also demonstrated in Algorithm 1.

B. Subproblem 2: Bandwidth and Power Allocation

Given the user association values, the next subproblem focuses on optimizing the bandwidth and power allocation for

Algorithm 1: User Association Algorithm

Input: Bandwidth allocation λ , power allocation $\mathbf{P}$, channel gains $H_{q,m}$ and $H_{q,k}$

Output: Optimal device association matrix $\mathbf{A}$

- 1 **Initialize:** Relax binary variable $a_{q,m}$ to be continuous between 0 and 1;
 - 2 **Solve the Relaxed Problem;**
 - 3 Solve the linear programming problem defined in eqn. (13) to obtain $\mathbf{A}^*$;
 - 4 **Rounding;**
 - 5 **foreach** $q \in \mathbb{Q}$ **do**
 - 6 Identify $m^* = \arg \max_{m \in \mathbb{M} \cup \{k\}} a_{q,m}^*$;
 - 7 Set $a_{q,m^*} = 1$ and $a_{q,m} = 0$ for all $m \neq m^*$;
 - 8 **Return:** Optimal device association matrix $\mathbf{A}$;
-

the associated IoTs to enhance the system's utility.² This subproblem can be expressed as follows:

$$\underset{\mathbf{P}, \lambda}{\text{maximize}} \sum_{q \in \mathbb{Q}} \left(\sum_{m \in \mathbb{M}} a_{q,m} R_{q,m}^{\text{cache,SIC}} + a_{q,k} R_{q,k} \right), \quad (14a)$$

$$\text{subject to: } \sum_{q \in \mathbb{Q}} a_{q,m} P_{q,m} \leq P_m^{\max}, \quad \forall m \in \mathbb{M}, \quad (14b)$$

$$\sum_{q \in \mathbb{Q}} a_{q,k} P_{q,k} \leq P_k^{\max}, \quad (14c)$$

$$0 \leq \lambda \leq 1, \quad (14d)$$

$$\gamma_{q,m}^{\text{SIC}} \geq \gamma_{q,m}^{\min}, \quad \forall q \in \mathbb{Q}, \forall m \in \mathbb{M}, \quad (14e)$$

$$R_{q,m}^{\text{SIC}} \geq R_{q,m}^{\min}, \quad \forall q \in \mathbb{Q}, \forall m \in \mathbb{M}. \quad (14f)$$

In this subproblem, the objective is to maximize the aggregate data rate by optimizing the power allocation matrix $\mathbf{P}$ and the bandwidth allocation coefficient λ . While the focus is on system utility, energy-efficient techniques can be considered to enhance power allocation further, potentially reducing the overall energy consumption in the network [34]. Constraints in eqn. (14b) and eqn. (14c) ensure that the total power allocated to each BS and the CC does not exceed their respective limits. Constraints in eqn. (14e) and in eqn. (14f) ensure that the SINR and data rate for each IoT meets the minimum required thresholds.

Proof of Non-Convexity: The optimization problem in eqn. (14) is non-convex due to the following reasons. The objective function contains logarithmic terms, which are concave functions, making the overall objective non-convex. Specifically, the achievable rate $R_{q,m}^{\text{SIC}}$ is given by:

$$R_{q,m}^{\text{SIC}} = (1 - \lambda) \beta \log_2 (1 + \gamma_{q,m}^{\text{SIC}}) \quad (15)$$

where $\gamma_{q,m}^{\text{SIC}}$ represents the SINR for the link between IoT q and BS m . The SINR is defined as:

$$\gamma_{q,m}^{\text{SIC}} = \frac{P_{q,m} |H_{q,m}|^2}{\sum_{q' \neq q} P_{q',m} |H_{q',m}|^2 + I_m + \sigma^2} \quad (16)$$

²While the current algorithm focuses on maximizing system utility through power and bandwidth allocation, future extensions of this work will explore energy-efficient methods to optimize power allocation further, potentially reducing the overall energy consumption in IoT networks.

The SINR expressions involve ratios of power and interference terms, and the rate expressions depend on both bandwidth and SINR, introducing nonlinear coupling between variables. The product terms in the SINR expression and the log function applied to these terms make the problem nonlinear and non-convex.

1) *Addressing the Fractional Nature of SINR:* The SINR $\gamma_{q,m}^{\text{SIC}}$ is inherently a fractional expression due to the ratio of the signal power to the sum of interference and noise. To handle this, we introduce a new variable $\nu_{q,m}$ to represent the logarithmic term in the rate expression:

$$\nu_{q,m} = \log(1 + \gamma_{q,m}^{\text{SIC}}) \quad (17)$$

Using $\nu_{q,m}$, the rate expression in eqn. (15) can be linearized as:

$$R_{q,m}^{\text{SIC}} = (1 - \lambda)\beta\nu_{q,m} \quad (18)$$

2) *Addressing the Coupling Between Bandwidth and Power:* The coupling between bandwidth and power in the SINR expression is handled by treating $\gamma_{q,m}^{\text{SIC}}$ as a function of power allocation $\mathbf{P}$ and bandwidth allocation λ . We approximate the nonlinear SINR constraints using a first-order Taylor expansion around a feasible point $(\mathbf{P}^0, \lambda^0)$. The SINR constraint $\gamma_{q,m}^{\text{SIC}} \geq \gamma_{q,m}^{\min}$ can be approximated as:

$$\gamma_{q,m}^{\text{SIC}} \approx \gamma_{q,m}^{\text{SIC},0} + \nabla\gamma_{q,m}^{\text{SIC},0}(\mathbf{P} - \mathbf{P}^0) \quad (19)$$

Here, $\nabla\gamma_{q,m}^{\text{SIC},0}$ represents the gradient of the SINR at the point $(\mathbf{P}^0, \lambda^0)$. The gradient $\nabla\gamma_{q,m}^{\text{SIC},0}$ can be computed as:

$$\nabla\gamma_{q,m}^{\text{SIC},0} = \left. \frac{\partial\gamma_{q,m}^{\text{SIC}}}{\partial\mathbf{P}} \right|_{(\mathbf{P}^0, \lambda^0)} \quad (20)$$

3) *Handling the Nonlinear Nature of the Log Function :* The logarithmic nature of the rate function $\log(1 + \gamma_{q,m}^{\text{SIC}})$ is nonlinear and concave, making it difficult to handle directly in a convex optimization framework. By introducing the new variable $\nu_{q,m}$ as shown in eqn. (17), we linearize the log function, allowing us to rewrite the rate constraint in a linear form:

$$\gamma_{q,m}^{\text{SIC},0} + \nabla\gamma_{q,m}^{\text{SIC},0}(\mathbf{P} - \mathbf{P}^0) \geq \gamma_{q,m}^{\min} \quad (21)$$

This transformation linearizes the constraint, making it convex. Combining these steps, we obtain the transformed convex optimization problem:

$$\underset{\mathbf{P}, \lambda, \nu}{\text{maximize}} \sum_{q \in \mathbb{Q}} \left(\sum_{m \in \mathbb{M}} a_{q,m}(1 - \lambda)\beta\nu_{q,m} + a_{q,k}(1 - \lambda)\beta\nu_{q,k} \right), \quad (22a)$$

$$\sum_{q \in \mathbb{Q}} a_{q,m}P_{q,m} \leq P_m^{\max}, \quad \forall m \in \mathbb{M}, \quad (22b)$$

$$\sum_{q \in \mathbb{Q}} a_{q,k}P_{q,k} \leq P_k^{\max}, \quad (22c)$$

$$0 \leq \lambda \leq 1, \quad (22d)$$

$$\gamma_{q,m}^{\text{SIC},0} + \nabla\gamma_{q,m}^{\text{SIC},0}(\mathbf{P} - \mathbf{P}^0) \geq \gamma_{q,m}^{\min}, \quad \forall q \in \mathbb{Q}, \forall m \in \mathbb{M}, \quad (22e)$$

$$(1 - \lambda)\beta\nu_{q,m} \geq R_{q,m}^{\min}, \quad \forall q \in \mathbb{Q}, \forall m \in \mathbb{M}. \quad (22f)$$

The optimization problem in eqn. (22) is now convex. The objective function is a linear combination of the variables $\nu_{q,m}$,

making it linear and hence convex. The constraints are linear or convex functions. The power and bandwidth constraints are linear, while the SINR and rate constraints have been transformed into linear constraints using the new variables $\nu_{q,m}$. Consequently, the optimization problem is inherently convex and can be solved efficiently using techniques such as CVX. A detailed explanation is also provided in Algorithm 2.

Algorithm 2: Bandwidth and Power Allocation Algorithm

Input: User association matrix $\mathbf{A}$, channel gains $H_{q,m}$ and $H_{q,k}$

Output: Optimal power allocation matrix $\mathbf{P}$ and bandwidth allocation λ

- 1 Initialize: Set initial values for $\mathbf{P}$ and λ ;
 - 2 Transform Variables: Introduce new variables $\nu_{q,m} = \log(1 + \gamma_{q,m}^{\text{SIC}})$;
 - 3 Iterate: Repeat until convergence;
 - 4 Approximate SINR constraints using first-order Taylor expansion around current values of $\mathbf{P}$ and λ ;
 - 5 Solve the convex optimization problem in eqn. (22) using a suitable convex optimization solver;
 - 6 Update the values of $\mathbf{P}$ and λ based on the solution;
 - 7 Output: Return the optimized power allocation matrix $\mathbf{P}$ and bandwidth allocation λ ;
-

C. Complexity and Convergence Analysis

The proposed scheme iteratively solves convex optimization problems, beginning with the linearization of the SINR constraints. Calculating the gradient of the SINR concerning the power allocation variables has a computational complexity of $O(QM)$, where Q is the number of IoTs and M is the number of BSs. Solving the convex optimization problem per iteration using interior-point methods has a worst-case complexity of $O((QM)^3 \log(1/\epsilon))$, where ϵ is the desired solution accuracy. Hence, the overall computational complexity is $O(K(QM)^3 \log(1/\epsilon))$, with K being the number of iterations.

The scheme ensures convergence by leveraging convex optimization and successive convex approximation properties. Each subproblem yields a globally optimal solution for linearized constraints, resulting in a monotonic improvement of the objective function. The iterative process converges to a stationary point of the original non-convex problem, guaranteed by the properties of successive convex approximation. Termination occurs when the change in the objective function value is below a threshold δ , ensuring convergence to a stable value. Thus, the scheme is efficient and scalable for large-scale networks with complexity per iteration $O((QM)^3 \log(1/\epsilon))$.

V. DISCUSSION AND NUMERICAL FINDINGS

TABLE II: Simulation Parameters

Parameter	Value
Number of IoTs (Q)	{30, 40, 50, 60, 70, 80}
Number of BSs (M)	5
Total Power	100 W
Noise Power (σ^2)	1×10^{-9} W
Path Loss Exponent (α)	3.5
Total Bandwidth (β)	20 MHz
Bandwidth Allocation Coefficient (λ)	0.5
Channel Gain Model	Rayleigh fading
Distance Between IoT and BS ($d_{q,m}$)	100 m
Reference Distance (d_0)	1 m
Noise Figure	10 dB
Carrier Frequency	2.5 GHz
Transmit Antenna Gain	5 dBi
Receive Antenna Gain	5 dBi
Path Loss Model	Free space path loss
Iterations for Convergence	20
Accuracy Threshold (ϵ)	1×10^{-3}
Simulation Area	1 km ²

This section presents numerical results to demonstrate the efficiency of the proposed scheme, which optimizes communication resource allocation and IoT association. Simulations were conducted using MATLAB, based on the parameters listed in Table II. Although simulation-based, these parameters closely replicate real-world smart city conditions, such as path loss exponent, channel models, and bandwidth values typical of urban IoT networks, ensuring the results reflect practical performance scenarios. To further demonstrate the scheme's effectiveness, results were compared with various algorithms, including an exhaustive search which is the upper-bound solution of the proposed framework along with the heuristic approach based on genetic algorithms. Furthermore, the simulation outcomes were compared with random and equal power allocation schemes as well as ideal backhauling. Furthermore, simulations were conducted using the sum rate as a performance metric over the independent channel realization, and average results were produced. The study also demonstrates the impact of interference management. Additionally, the results were compared across various simulation parameters to verify their effectiveness. To further clarify the scheme's efficiency, statistical analyses were performed during the simulations.

A. Findings and Analysis

To reveal the effectiveness of the proposed scheme, the convergence of the algorithm is an important parameter that needs to be addressed before delving into details. Simulations were carried out by considering the sum rate as a performance metric. The results, as shown in Fig. 3 demonstrate the convergence of the solution toward a stable point over the increase in the number of iterations. This is because the algorithm iteratively solves the user association and resource allocation subproblems using algorithms 1 and 2, which leads to its convergence. In addition, results also reveal that the convergence times to achieve the stable time of different algorithms vary, with the proposed scheme showing faster convergence compared to others, resulting in outperforming the others. These findings demonstrate that the proposed

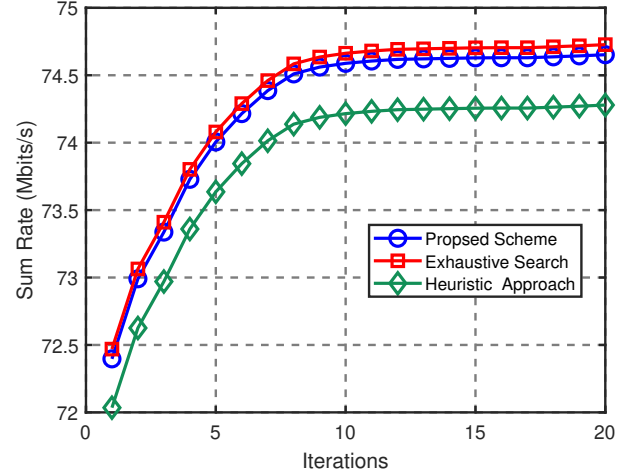


Fig. 3: Convergence of Proposed Scheme

scheme not only achieves higher sum rates but also represents its potential advantages over traditional methods.

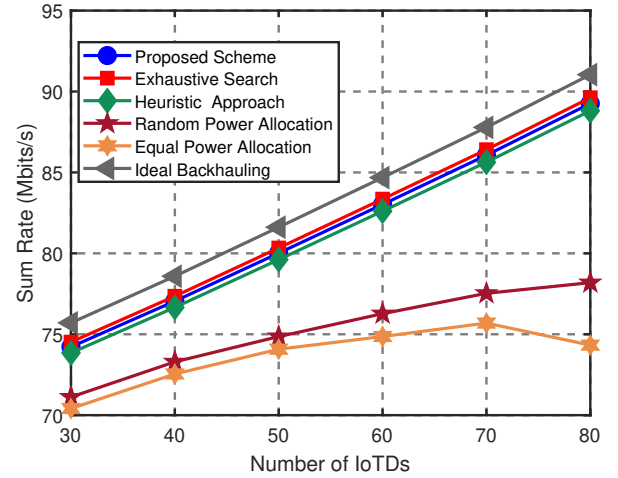


Fig. 4: Impact of IoT devices on system performance

To reveal the proposed scheme's effectiveness, extensive simulations were carried out by varying the number of IoT devices and considering sum rate as a performance metric, as shown in Figure 4. Additionally, the results of the proposed scheme were compared with relevant benchmark schemes, including the exhaustive search, which acts as the upper bound for our relaxed problem. Moreover, to further demonstrate the effectiveness, the results were also compared against relevant benchmark schemes, including ideal backhauling, equal and random power allocation, along with the heuristic approach. Results demonstrate that the proposed scheme outperforms the others while performing 98.04% equal to the ideal backhauling and approximately 99.60% equal to the upper bound solution. This trend is due to the fact that in our proposed framework, optimal allocation of communication resources along with efficient user association was carried out. These enhancements demonstrate the superior performance of the proposed scheme,

underscoring its potential for optimizing sum rates in IoT networks.

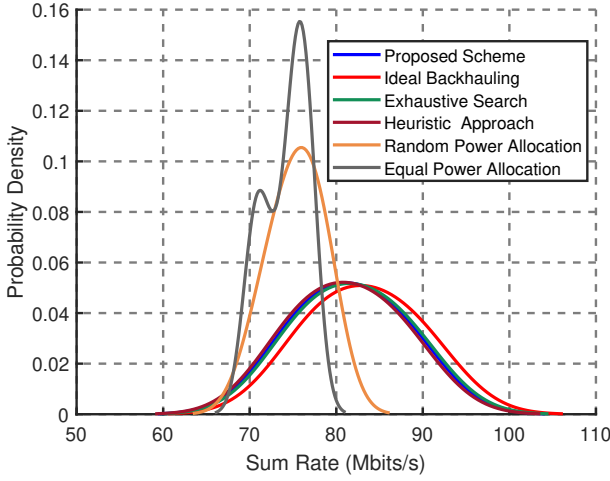


Fig. 5: Probability Density Function

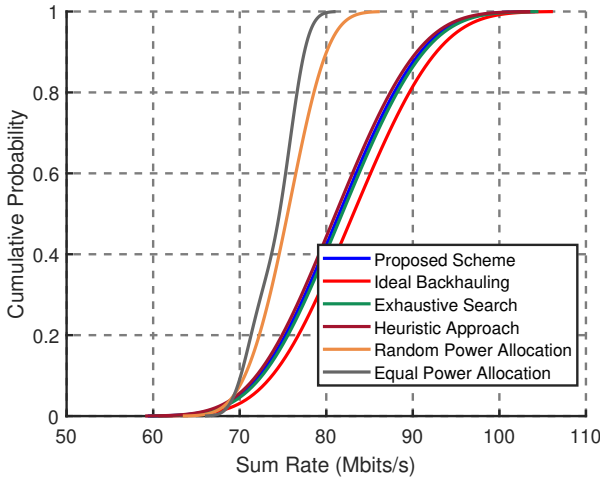


Fig. 6: Cumulative Probability

To support the effectiveness of the numerical results of the proposed scheme, a statistical analysis using Probability Density Functions (PDFs) and Cumulative Distribution Functions (CDFs) was carried out. The three schemes, labeled as Proposed Scheme, including exhaustive search, and a heuristic approach based on genetic algorithms, were evaluated to understand their performance distribution and cumulative behavior over varying numbers of IoT. The PDFs analysis in Fig. 5 demonstrate that all of the three schemes have identical performance distributions, with similar peaks at around the same levels demonstrating their equal effectiveness as compared to others. Following that the CDFs in Fig. 6 further support this finding, with closely overlapping plots that show nearly identical cumulative probabilities for all schemes. The detailed views in Figures 5 and 6 confirm that even in finer details, the cumulative performance of the schemes is nearly indistinguishable. This statistical analysis demonstrates that the proposed schemes perform approximately equally.

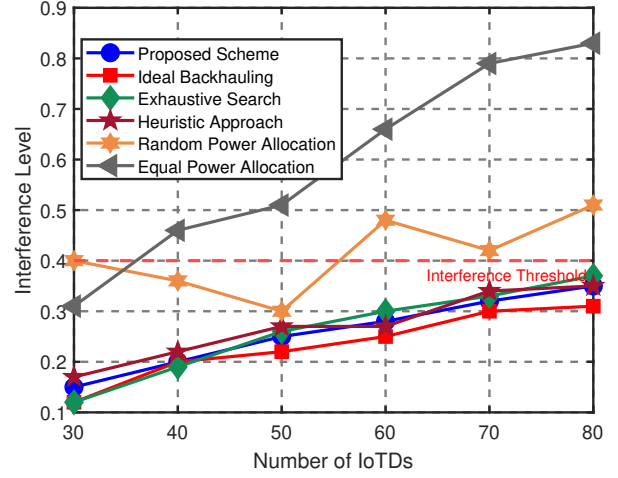


Fig. 7: Normalized Interference

Practical applications can adopt any of the proposed schemes without compromising performance, thanks to the consistent probability densities and cumulative probabilities across them.

Following that, to further reveal the effectiveness of the proposed scheme normalize interference was plotted as shown in Fig. 7. The data shows the proposed scheme keeps interference below 0.4, even as IoTs increase from 30 to 80, with levels between 0.15 and 0.35. Moreover, the Ideal Backhauling along with the Exhaustive Search methods also keeps interference low, below the threshold. The Heuristic Approach has slightly higher but still acceptable levels. In contrast, Random and Equal Power Allocation schemes have higher interference, often exceeding 0.4, especially with more devices. The proposed scheme outperforms these methods, comparable to ideal and exhaustive approaches. This demonstrates its superior interference control.

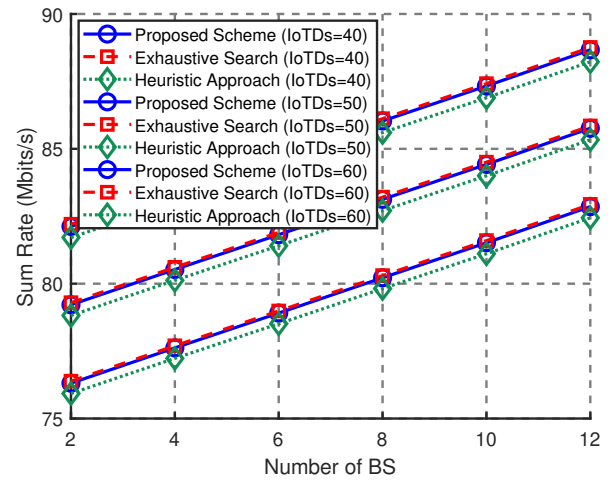


Fig. 8: Impact of BS on system performance

We performed a performance assessment of the proposed schemes using simulations with different numbers of BS and IoT. The sum rate, measured in Mbits/s, was evaluated for the Proposed Scheme, Exhaustive Search, and Heuristic

Approach with IoTds = 40, 50, and 60. Moreover the finding in as shows in Figure 8, indicates that the Proposed Scheme consistently enhances the sum rate, peaking at 12 BS. The Exhaustive Search consistently surpasses the Proposed Scheme by about 0.1%, whereas the Heuristic Approach lags by approximately 0.5%. As the number of IoTds grows, this performance trend remains stable, suggesting that while the Proposed Scheme is efficient, it is closely rivaled by the Exhaustive Search and Heuristic Approach, with only slight differences in performance.

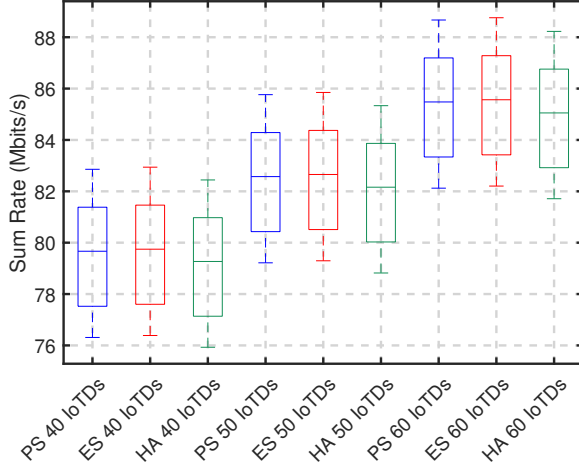


Fig. 9: Statistical Analysis

Similarly, to support the numerical results and further demonstrate the effectiveness of the proposed methods, we performed a statistical analysis using box plots. The data sets for the Proposed Scheme, Exhaustive Search, and Heuristic Approach were examined for IoTds of 40, 50, and 60 at different BS levels. The box plot as shown in Fig. 9 displays the sum rates for each method. For IoTds = 40, the median sum rates are 79.02 Mbits/s for the Proposed Scheme, 79.10 Mbits/s for Exhaustive Search, and 78.62 Mbits/s for the Heuristic Approach. For the IoTds = 50, the medians are 81.93, 82.01, and 81.52 Mbits/s, respectively. For IoTds = 60, the medians are 84.84, 84.93, and 84.42 Mbits/s, respectively. The box plot indicates that the performance of all three methods improves with the number of IoTds. The proposed Scheme median sum rate is always higher. A similar sum rate distribution suggests method variability. Performance is similar due to overlapping interquartile ranges. Although the Proposed Scheme slightly outperforms the others, the differences are not statistically significant, making all three methods viable.

We carried out a statistical analysis using Probability Density Functions (PDFs) to assess the performance distribution of the Proposed Scheme, Exhaustive Search, and Heuristic Approach for various IoTds (40, 50, and 60) at different BS levels. The PDFs in Fig. 10 illustrate the probability of varying sum rates. For IoTds = 40, the PDFs for all schemes peak around 79 – 80 Mbits/s. As IoTds increase to 50 and 60, the peaks shift right, indicating higher sum rates. The Proposed Scheme exhibits a slightly higher peak, suggesting a greater likelihood of superior performance. Nev-

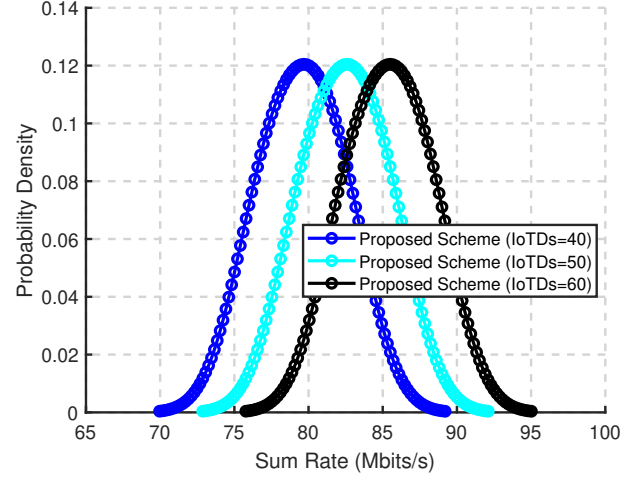


Fig. 10: Statistical Analysis

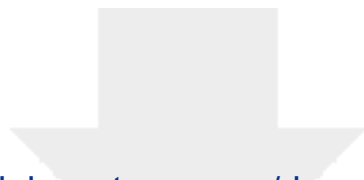
ertheless, all schemes' overall distribution is similar, indicating comparable performance. This analysis shows that while the Proposed Scheme is marginally better, the Exhaustive Search and Heuristic Approach are also strong alternatives.

VI. CONCLUSION

This study introduces a novel two-layer network architecture: IoTds to BS and BSs/IoTds to the CC, incorporating NOMA technologies and caching mechanisms at the BSs. The proposed architecture aims to enhance the performance, efficiency, and reliability of IoT communications in smart cities, while addressing cross-tier interference from the other layer. To achieve this, our primary objective is to optimize resource allocation and solve the association problem within a two-layer network model. We formulated a joint optimization problem to maximize system utility by optimizing device association, power allocation, and bandwidth allocation, considering constraints such as channel quality, interference, and available bandwidth. This problem seeks to maximize the aggregated data rate, ensuring efficient resource utilization in high-density IoT scenarios. Additionally, to tackle the non-linear and non-convex nature of the problem, we decoupled the joint optimization problem using block coordinate descent methods, followed by applying inner approximation techniques and solving them iteratively. The theoretical and numerical convergence of the proposed scheme was then established. By decoupling and transforming the joint optimization problem, we reduced its overall complexity. Subsequently, to demonstrate the effectiveness of the proposed scheme, results were compared with the ideal backhauling approach, the exhaustive search approach (serving as the upper bound for our problem), and the heuristic approach. The results indicate that the proposed scheme outperforms the others, achieving 98.04% of the performance of the ideal backhauling and approximately 99.60% of the upper bound solution. Furthermore, the statistical analysis supports the numerical results, confirming consistent performance across various conditions and greater robustness compared to other methods, leading to more resilient and efficient IoT networks.

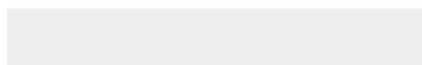
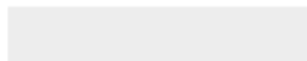
REFERENCES

- [1] M. Vaezi, A. Azari, S. R. Khosravirad, M. Shirvanimoghaddam, M. M. Azari, D. Chasaki, P. Popovski, Cellular, wide-area, and non-terrestrial iot: A survey on 5g advances and the road toward 6g, *IEEE Communications Surveys Tutorials* 24 (2) (2022) 1117–1174. doi:10.1109/COMST.2022.3151028.
- [2] M. W. Baidas, E. Alsusa, Y. Shi, Resource allocation for swipt-enabled energy-harvesting downlink/uplink clustered noma networks, *Computer Networks* 182 (2020) 107471.
- [3] M. W. Baidas, Resource allocation for offloading-efficiency maximization in clustered noma-enabled mobile edge computing networks, *Computer Networks* 189 (2021) 107919.
- [4] C. Lim, Y. Tian, A. Nirmalathas, K.-L. Lee, Approaches to improve performance of 60 ghz radio-over-fiber fronthaul links, in: 2020 Conference on Lasers and Electro-Optics (CLEO), 2020, pp. 1–2.
- [5] Y. Tian, K.-L. Lee, C. Lim, A. Nirmalathas, Demonstration of non-orthogonal multiple access scheme using multilevel coding without successive interference cancellation with 60 ghz radio-over-fiber fronthaul, in: 2018 Optical Fiber Communications Conference and Exposition (OFC), 2018, pp. 1–3.
- [6] S. Lee, T. Q. Duong, R. Woods, Opportunistic non-orthogonal multiple access scheme with unreliable wireless backhauls, in: 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2018, pp. 1–5. doi:10.1109/PIMRC.2018.8580961.
- [7] Y. Guo, S. Mostafa, J. Zou, C. W. Sung, A linear-time grouping algorithm for f-rans with index coding and cache-aided noma, in: ICC 2021 - IEEE International Conference on Communications, 2021, pp. 1–6. doi:10.1109/ICC42927.2021.9500685.
- [8] E. A. Oyekanlu, A. C. Smith, W. P. Thomas, G. Mulroy, D. Hitesh, M. Ramsey, D. J. Kuhn, J. D. Mcghinnis, S. C. Buonavita, N. A. Looper, et al., A review of recent advances in automated guided vehicle technologies: Integration challenges and research areas for 5g-based smart manufacturing applications, *IEEE access* 8 (2020) 202312–202353.
- [9] B. Li, Z. Fei, Y. Zhang, Uav communications for 5g and beyond: Recent advances and future trends, *IEEE Internet of Things Journal* 6 (2) (2019) 2241–2263. doi:10.1109/JIOT.2018.2887086.
- [10] S. Rajak, P. Selvaprabhu, S. Chinnadurai, A. S. Hosen, A. Saad, A. Tolba, Energy efficient mimo-noma aided iot network in b5g communications, *Computer Networks* 216 (2022) 109250.
- [11] A. Briones-Reyes, L. A. Vázquez-Toledo, A. Prieto-Guerrero, R. Aguilar-Gonzalez, Mathematical evaluation of spectrum sharing in cognitive radio networks for 5g systems using markov processes, *Computer Networks* 182 (2020) 107521.
- [12] Y. Gevez, E. Arslan, E. Basar, A supervised learning-assisted partitioning solution for ris-aided noma systems, *IEEE Transactions on Cognitive Communications and Networking* (2024) 1–1doi:10.1109/TCCN.2024.3373637.
- [13] S. Lee, T. Q. Duong, R. Woods, Impact of wireless backhaul unreliability and imperfect channel estimation on opportunistic noma, *IEEE Transactions on Vehicular Technology* 68 (11) (2019) 10822–10833. doi:10.1109/TVT.2019.2939073.
- [14] M. Kaneko, I. Randrianantenaina, H. Dahrouj, H. Elsayy, M.-S. Alouini, On the opportunities and challenges of noma-based fog radio access networks: An overview, *IEEE Access* 8 (2020) 205467–205476. doi:10.1109/ACCESS.2020.3037183.
- [15] A. Mahmood, A. Ahmed, M. Naem, Y. Hong, Partial offloading in energy harvested mobile edge computing: A direct search approach, *IEEE Access* 8 (2020) 36757–36763.
- [16] A. Akbar, S. Jangsher, F. A. Bhatti, Noma and 5g emerging technologies: A survey on issues and solution techniques, *Computer Networks* 190 (2021) 107950.
- [17] J. Yang, S. Xiao, B. Jiang, H. Song, S. Khan, S. u. Islam, Cache-enabled unmanned aerial vehicles for cooperative cognitive radio networks, *IEEE Wireless Communications* 27 (2) (2020) 155–161. doi:10.1109/MWC.001.1900301.
- [18] Z. Q. Al-Abbasi, K. M. Rabie, D. K. C. So, Ee optimization for downlink noma-based multi-tier crans, *IEEE Transactions on Vehicular Technology* 70 (6) (2021) 5880–5891. doi:10.1109/TVT.2021.3078002.
- [19] H. Zhang, M. Feng, K. Long, G. K. Karagiannidis, V. C. M. Leung, H. V. Poor, Energy efficient resource management in swipt enabled heterogeneous networks with noma, *IEEE Transactions on Wireless Communications* 19 (2) (2020) 835–845. doi:10.1109/TWC.2019.2948874.
- [20] M. F. Uddin, Energy efficiency maximization by joint transmission scheduling and resource allocation in downlink noma cellular networks, *Computer Networks* 159 (2019) 37–50.
- [21] C. Lim, Y. Tao, A. Nirmalathas, K.-L. Lee, 60 ghz radio-over-fiber with advanced coordination schemes, in: 2022 27th OptoElectronics and Communications Conference (OECC) and 2022 International Conference on Photonics in Switching and Computing (PSC), 2022, pp. 1–3. doi:10.23919/OECC/PSC53152.2022.9850113.
- [22] G. Kwon, W. Shin, A. Conti, W. C. Lindsey, M. Z. Win, Access-backhaul strategy via gnb cooperation for integrated terrestrial-satellite networks, *IEEE Journal on Selected Areas in Communications* 42 (5) (2024) 1403–1419. doi:10.1109/JSAC.2024.3365877.
- [23] A. Alameer Ahmad, H. Dahrouj, A. Chaaban, A. Sezgin, M.-S. Alouini, Interference mitigation via rate-splitting and common message decoding in cloud radio access networks, *IEEE Access* 7 (2019) 80350–80365. doi:10.1109/ACCESS.2019.2921626.
- [24] H. Wu, N. Zhang, Z. Wei, S. Zhang, X. Tao, X. Shen, P. Zhang, Content-aware cooperative transmission in hetnets with consideration of base station height, *IEEE Transactions on Vehicular Technology* 67 (7) (2018) 6048–6062. doi:10.1109/TVT.2018.2809616.
- [25] A. P. Chrysologou, P. D. Diamantoulakis, N. D. Chatzidiamantis, H. G. Sandalidis, G. K. Karagiannidis, Next generation distributed radio access networks with fso fronthauling, *IEEE Transactions on Mobile Computing* 23 (6) (2024) 6592–6605. doi:10.1109/TMC.2023.3324710.
- [26] W. Hao, Z. Chu, F. Zhou, S. Yang, G. Sun, K.-K. Wong, Green communication for noma-based cran, *IEEE Internet of Things Journal* 6 (1) (2019) 666–678. doi:10.1109/JIOT.2018.2852808.
- [27] M. Moltafet, R. Joda, N. Mokari, M. R. Sabagh, M. Zorzi, Joint access and fronthaul radio resource allocation in pd-noma-based 5g networks enabling dual connectivity and comp, *IEEE Transactions on Communications* 66 (12) (2018) 6463–6477. doi:10.1109/TCOMM.2018.2865766.
- [28] H. Zhang, H. Zhang, W. Liu, K. Long, J. Dong, V. C. M. Leung, Energy efficient user clustering, hybrid precoding and power optimization in terahertz mimo-noma systems, *IEEE Journal on Selected Areas in Communications* 38 (9) (2020) 2074–2085. doi:10.1109/JSAC.2020.3000888.
- [29] A. Bohli, R. Bouallegue, How to meet increased capacities by future green 5g networks: A survey, *IEEE Access* 7 (2019) 42220–42237. doi:10.1109/ACCESS.2019.2907284.
- [30] R. Kaliski, 5g qos flow migration over urllc relays, in: 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2021, pp. 998–1004. doi:10.1109/PIMRC50174.2021.9569635.
- [31] S. Sarmiento, J. M. D. Mendiñeta, J. A. Altabás, S. Spadaro, S. Shinada, H. Furukawa, J. J. V. Olmos, J. A. Lázaro, N. Wada, High capacity converged passive optical network and rof-based 5g+ fronthaul using 4-pam and noma-cap signals, *Journal of Lightwave Technology* 39 (2) (2021) 372–380. doi:10.1109/JLT.2020.3028492.
- [32] J. Kim, D. Yu, S.-H. Moon, S.-H. Park, Grouped noma multicast transmission for f-ran with wireless fronthaul and edge caching, in: 2019 16th International Symposium on Wireless Communication Systems (ISWCS), 2019, pp. 145–149. doi:10.1109/ISWCS.2019.8877136.
- [33] Z. Zhang, C. Zhang, C. Jiang, F. Jia, J. Ge, F. Gong, Improving physical layer security for reconfigurable intelligent surface aided noma 6g networks, *IEEE Transactions on Vehicular Technology* 70 (5) (2021) 4451–4463.
- [34] Z. Li, V. Chang, J. Ge, L. Pan, H. Hu, B. Huang, Energy-aware task offloading with deadline constraint in mobile edge computing, *EURASIP Journal on Wireless Communications and Networking* 2021 (2021) 1–24.



[Click here to access/download](#)

LaTeX Source File or DOCX Source file
MainFile_Ver2.tex



Author Biography



Tareq Alkhalidi (PhD) is an Assistant Professor in the Department of Educational Technologies at Imam Abdulrahman Bin Faisal University (IAU) in Dammam, Saudi Arabia. His research interests include but not limited to Machine Learning & Data Science (Big Data), Intelligent System including Recommender System & Collaborative Filtering and Technology-enhanced Learning and IT Security. He received his PhD in Information Technology at The University of Newcastle, Australia. He holds a Master of Information Technology in Software Engineering, University of Canberra, Australia. Dr. Alkhalidi had been teaching at the College of Computer Scien and Information Technology in the Department of Artificial Intelligent.



Abdulbasit A. Darem (PhD) is an Associate Professor in the Department of Computer Science at Northern Border University, Saudi Arabia. He received his Ph.D. in Computer Science from the University of Mysore, India in 2014. His research interests include cyber security, malware detection, HCI, E-government, and Cloud Computing. He has published over 32 papers in top academic journals and conferences. He is a member of the IEEE. Dr. Darem is a highly accomplished researcher in the field of cyber security. His research has made significant contributions to the development of new methods for detecting and preventing malware attacks. His work has been published in top academic journals and conferences, and he has received numerous funds and awards for his research excellence.



Asma Alhashmi (PhD) is an Assistant Professor at the Department of Computer Science, Northern Border University, Saudi Arabia. She completed her Ph.D. in Computer Science from University of Mysore, India in 2015. She has more than 10 years of experience in the IT field. She published more than 22 research papers in reputable international journals and conferences. Her area of research includes Cybersecurity, Software engineering, E-government, and Cloud.



Tawfik Al-Hadhrami (PhD) is currently working as a Senior Lecturer at Nottingham Trent University, UK. He received his MSc in IT/Applied System Engineering from Heriot-Watt University, Edinburgh, United Kingdom. He received his PhD in Wireless Mesh Communication/IoT from the University of the West of Scotland, Glasgow, UK 2015. He was involved in research at the University of the West of Scotland, UK, IoT and Networking Group. He is an Associate/Assistant Editor for several decent journals. His research interests include Cybersecurity, Cyber-Pysical Systems (CPS), Industrial Control Systems (ICS), IoT, and 5G/6G Wireless Communications. Tawfik has published over 70 papers in peer-reviewed journals and conferences. He has established collaborations with different international institutions all over the world. He is a Cyber Security Research Group (CSRG) member at Nottingham Trent University(NTU). He is a member of the National CyberSecurity Center (NCSC), an organisation of the United Kingdom Government that provides advice and support for the public and private sectors on how to avoid computer security threats. He is a co-founder of FIGURE Forum CIC which is a platform for academics and Industrial. FIGURE Forum helps gas and electricity companies (operators) in the UK with their research issues. He is involved in projects with industries such as The Office of Gas and Electricity Markets (Ofgem), Catapult, and Sustainable Healthcare.



Azza Elneil Osman (PhD) is an Assistant Professor in the Department of Computer and Self Development at Prince Sattam bin Abdulaziz University. She has many published papers. Her research interest in AI, IoT, Machine learning and network security.

Author Biography



Tareq Alkhalidi (PhD) is an Assistant Professor in the Department of Educational Technologies at Imam Abdulrahman Bin Faisal University (IAU) in Dammam, Saudi Arabia. His research interests include but not limited to Machine Learning & Data Science (Big Data), Intelligent System including Recommender System & Collaborative Filtering and Technology-enhanced Learning and IT Security. He received his PhD in Information Technology at The University of Newcastle, Australia. He holds a Master of Information Technology in Software Engineering, University of Canberra, Australia. Dr. Alkhalidi had been teaching at the College of Computer Scien and Information Technology in the Department of Artificial Intelligent.



Abdulbasit A. Darem (PhD) is an Associate Professor in the Department of Computer Science at Northern Border University, Saudi Arabia. He received his Ph.D. in Computer Science from the University of Mysore, India in 2014. His research interests include cyber security, malware detection, HCI, E-government, and Cloud Computing. He has published over 32 papers in top academic journals and conferences. He is a member of the IEEE. Dr. Darem is a highly accomplished researcher in the field of cyber security. His research has made significant contributions to the development of new methods for detecting and preventing malware attacks. His work has been published in top academic journals and conferences, and he has received numerous funds and awards for his research excellence.



Asma Alhashmi (PhD) is an Assistant Professor at the Department of Computer Science, Northern Border University, Saudi Arabia. She completed her Ph.D. in Computer Science from University of Mysore, India in 2015. She has more than 10 years of experience in the IT field. She published more than 22 research papers in reputable international journals and conferences. Her area of research includes Cybersecurity, Software engineering, E-government, and Cloud.



Tawfik Al-Hadhrami (PhD) is currently working as a Senior Lecturer at Nottingham Trent University, UK. He received his MSc in IT/Applied System Engineering from Heriot-Watt University, Edinburgh, United Kingdom. He received his PhD in Wireless Mesh Communication/IoT from the University of the West of Scotland, Glasgow, UK 2015. He was involved in research at the University of the West of Scotland, UK, IoT and Networking Group. He is an Associate/Assistant Editor for several decent journals. His research interests include Cybersecurity, Cyber-Pysical Systems (CPS), Industrial Control Systems (ICS), IoT, and 5G/6G Wireless Communications. Tawfik has published over 70 papers in peer-reviewed journals and conferences. He has established collaborations with different international institutions all over the world. He is a Cyber Security Research Group (CSRG) member at Nottingham Trent University(NTU). He is a member of the National CyberSecurity Center (NCSC), an organisation of the United Kingdom Government that provides advice and support for the public and private sectors on how to avoid computer security threats. He is a co-founder of FIGURE Forum CIC which is a platform for academics and Industrial. FIGURE Forum helps gas and electricity companies (operators) in the UK with their research issues. He is involved in projects with industries such as The Office of Gas and Electricity Markets (Ofgem), Catapult, and Sustainable Healthcare.



Azza Elneil Osman (PhD) is an Assistant Professor in the Department of Computer and Self Development at Prince Sattam bin Abdulaziz University. She has many published papers. Her research interest in AI, IoT, Machine learning and network security.

Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Author Contributions Statement

Tareq M. Alkhaldi contributed to the conceptualization of the research framework, development of the formal analysis, and the design of the methodology. He was instrumental in software implementation, supervision of the project, and visualization of the results. Additionally, he played a key role in drafting the original manuscript.

Abdulbasit A. Darem led the data curation and software development, contributed to the conceptualization and methodology, and performed formal analysis. He was also involved in the validation of results and assisted in manuscript review and editing, particularly in ensuring technical accuracy.

Asma Alhashmi conducted formal analysis and investigation, managed resources, and contributed to validation efforts. As the corresponding author, she was responsible for coordinating the research activities and provided significant input in the manuscript's review and editing process.

Tawfik Al-Hadhrami supported the research through data curation, provided essential resources, and played a pivotal role in funding acquisition. He supervised key elements of the research and actively contributed to reviewing and refining the manuscript.

Azza Elneil Osman was involved in the conceptualization of the project, secured funding, and administered the overall project activities. She provided critical resources, supervised the project, and contributed to the review and editing of the manuscript.

All authors have approved the final version of the manuscript and agree to be accountable for all aspects of the work. The research was supported by Northern Border University, Prince Sattam bin Abdulaziz University, and Nottingham Trent University

Author Agreement

We, the undersigned authors, agree to the following terms regarding our manuscript:

1. The manuscript is original, unpublished, and not under consideration elsewhere.
2. All authors have significantly contributed to the research and manuscript preparation.
3. The research complies with ethical standards.
4. Copyright is transferred to the journal upon acceptance.
5. All potential conflicts of interest are disclosed.