
R. Porselvi1*, D. Kanchana2, Beulah Jackson3, L. Vigneash4

Abstract
Given the burgeoning demands of real-time applications and limited on-board computational resources, the rapid proliferation of 6G-enabled vehicular networks brings forth the critical challenge of efficient computation offloading. This paper introduces the CORA-6G algorithm, a novel solution specifically designed to address this issue by determining the optimal offloading ratio. Using a robust convex optimization problem, CORA-6G maximizes the computational utility, ensuring efficient resource allocation and enhanced responsiveness. Quantitative evaluations reveal that, when compared to conventional methods, CORA-6G improves offloading efficiency by up to 25%, thereby significantly boosting the overall performance of vehicular applications. This research underscores the potential of CORA-6G as a cornerstone for future 6G vehicular network advancements.

Keywords: 6G, Vehicular network, Offloading, Convex optimization, Resource allocation.

Introduction
Empowered by 5G capabilities, vehicular networks are forging connections among billions of vehicles. These networks serve as a foundation for pioneering applications like interactive gaming, facial recognition, and augmented reality. To captivate potential users, mobile apps within this space are often dynamic and feature-rich. However, this richness demands significant computational power and drains battery energy, posing a substantial challenge for facilities with limited resources (Anitha G. et al., 2022).

The increasing complexity and computational demands of applications in vehicular networks necessitate a more sophisticated approach to resource management (Ayyadurai M. et al., 2022). Traditional local computation often falls short in meeting these demands, leading to increased latency and rapid battery depletion. This is where the need for a computation offloading strategy becomes imperative. Offloading tasks to more powerful external computational resources, such as edge or cloud servers, can significantly reduce the computational burden on local devices, thereby decreasing latency and conserving battery life. As the next evolutionary step, 6G promises even higher data rates and lower latencies but will also bring forth more computationally intensive applications (Boobalan S. et al., 2021).

Various approaches to computation offloading and resource allocation have been explored in the literature, each addressing different facets of the problem. A framework that allows a single user to offload tasks to multiple servers is presented in Cui Y. et al., (2016), with the optimization problem formulated to minimize both energy consumption and task execution latency. Dynamic resource and task allocation are the focus of the study in Di B. et al., (2016), which aims to minimize the network's total energy cost by regulating local CPU and cloud resources. In Dinh T. Q. et al., (2017), a comprehensive approach is taken that jointly considers computation offloading, resource allocation, and content caching in heterogeneous wireless networks.
proposing a dynamic method to maximize total network revenue. An adaptive online scheduling algorithm designed to improve mobile devices’ energy efficiency is introduced in Jošilo S. & Dán, G. (2018), targeting the minimization of the total energy cost for multiple mobile applications while meeting user performance expectations. The work in Kwak J. et al., (2015) integrates cloud radio networks and mobile edge computing (MEC) to explore dynamic resource scheduling, aiming to maximize the service provider’s profit by balancing power consumption and performance through an improved Lyapunov algorithm. Similarly, the study in Liu Y. et al., (2018) investigates resource scheduling with partial information, employing a perturbed Lyapunov technique to maximize network utility by adapting to historical network knowledge.

Existing offloading approaches often suffer from limitations such as limited scalability, as they are primarily designed for single-user scenarios. While some methods focus on energy efficiency, they may not adapt well to dynamic network conditions or could compromise on other key performance metrics like latency. Other more comprehensive solutions are too complex for practical implementation or do not fully leverage the capabilities of emerging technologies like 6G. Additionally, some approaches rely on potentially outdated network information, which could lead to suboptimal resource allocation decisions.

The CORA-6G algorithm proposed in this paper aims to address these challenges by offering a more scalable, adaptive, and comprehensive solution for computation offloading and resource allocation in next-generation vehicular networks. The contributions of this paper are as follows:

- Presentation of the CORA-6G algorithm optimizes both latency and energy consumption while ensuring efficient resource allocation and offloading in large-scale vehicular networks.
- An extensive performance evaluation of the CORA-6G algorithm, including scalability tests with varying numbers of IoT devices to assess its efficiency and robustness.

The remainder of this paper is structured as follows: Section II reviews existing work in computation offloading and resource allocation, setting the stage for the contributions of this study. Section III delves into the details of the proposed CORA-6G algorithm, explaining its components and functionalities. Section IV includes both the experimental setup and the performance evaluation of CORA-6G, along with a discussion on the results and their implications for 6G vehicular networks. Finally, section V concludes the paper and highlights potential directions for future research in this area.

**Related Works**

Numerous pivotal studies have been conducted in the field of computation offloading and resource allocation. A representative work by (Lyu, X., et al., 2017) set the stage by concentrating on energy-efficient resource allocation in multiuser MEC offloading systems, employing convex optimization to minimize energy consumption. This seminal work paved the way for subsequent research by (Ning, Z., et al., 2020), who broadened the scope to include distributed cloud and edge computing systems, with an objective of maximizing overall system profit while accounting for multiple types of resources. Expanding upon these foundational studies, (Oo, T. Z., et al., 2016) utilized deep reinforcement learning (DRL) paradigm to jointly optimize computation offloading and resource allocation in MEC. In this context (Seid, A. M., et al., 2021) introduced a novel two-stage approach using DRL for computation offloading and resource allocation in multi-UAV-assisted IoT networks, specifically targeting emergency scenarios. These contributions collectively underscore the progressive advancements in strategies for optimizing resource allocation and computation offloading across diverse computing landscapes.

Beginning with (Wang C. et al., 2017) exploration into the intricacies of non-orthogonal multiple access (NOMA) networks, the optimization of sub-channel assignment, power allocation, and user scheduling has been put forth as a method to capitalize on the spectral efficiencies offered by NOMA. Venturing into the realm of computation offloading, (Wang X. et al., 2020) extended the principles of NOMA to design hybrid offloading strategies in fog and cloud networks. Their work aptly demonstrates how integrating NOMA with computational offloading can be pivotal, especially in resource-intensive scenarios, ensuring timely data processing and efficient resource usage. To further refine the offloading process (Wang X. et al., 2018) proposed a decentralized, game-theoretic approach. Unlike centralized methods, where decisions are made centrally, this approach lets individual devices in a mobile cloud computing environment make independent offloading decisions. This autonomy ensures that each device’s unique computational needs and resources are adequately addressed, thereby optimizing the overall system’s efficiency.

(Yuan, H., & Zhou, M. 2020) present partial computation offloading and adaptive task scheduling algorithm (POETS), aimed at optimizing total system profit. Utilizing a two-sided matching algorithm, it achieves optimal transmission scheduling, while convex optimization is employed to ascertain individual offloading ratios without the need for cross-user data. Additionally, a non-cooperative game is constructed to establish a balance between user interests and network operator objectives. The algorithm's efficacy is confirmed through both theoretical analysis and empirical tests based on real-world data.

While existing studies have made strides in either computation offloading or resource allocation, a research gap exists in addressing both elements in an integrated manner. Current works often focus on specific network conditions or user behaviors, limiting their applicability and generalizability. This segmented approach can result in suboptimal network performance, as offloading and resource allocation are intrinsically linked. The lack of comprehensive solutions that combine these two critical aspects, especially in emerging technologies like 6G vehicular networks, underscores the need for a unified approach to optimize individual metrics and overall network efficiency.

**CORA-6G System Architecture**

In this section, an overview of the CORA-6G algorithm is provided. Designed to address the challenges of computation offloading and resource allocation in emerging 6G vehicular networks, CORA-6G aims to optimize system utility while adhering to latency and energy constraints.

**Problem formulation**

This subsection formally defines the problem of computation offloading and resource allocation in 6G vehicular networks. Let \( N \) be the set of vehicular users, and \( M \) be the set of available edge servers. The objective is to maximize the overall system utility \( U \) while minimizing the latency \( L \) and energy consumption \( E \). The objective function is defined as in (1), where \( \alpha, \beta \) and \( \gamma \) are weighting factors, \( U_{\text{comp},i} \) is the computation utility for user \( i \), and \( U_{\text{trans},i} \) is the transmission utility for user \( i \).

\[
U = \sum_{i \in N} \left( \alpha \cdot U_{\text{comp},i} + \beta \cdot U_{\text{trans},i} \right) - \gamma \cdot (L + E) \quad (1)
\]

The latency \( L \) is given as in (2), where \( L_{\text{comp},i} \) is the computation latency and \( L_{\text{trans},i} \) is the transmission latency for user \( i \).

\[
L = \sum_{i \in N} (L_{\text{comp},i} + L_{\text{trans},i}) \quad (2)
\]

The energy consumption \( E \) is defined as in (3), where \( E_{\text{comp},i} \) is the computation energy and \( E_{\text{trans},i} \) is the transmission energy for user \( i \).

\[
E = \sum_{i \in N} (E_{\text{comp},i} + E_{\text{trans},i}) \quad (3)
\]

The optimization problem can be stated as below to find the optimal computation offloading and resource allocation strategy that maximizes \( U \) while satisfying the given constraints.

\[
\begin{align*}
\text{Maximize} \quad & U \\
\text{Subject to} \quad & C_{\text{comp},i} \leq C_{\text{max},i}, \forall i \in N \\
& B_{\text{trans},i} \leq B_{\text{max},i}, \forall i \in N \\
& E \leq E_{\text{max}}
\end{align*}
\]

- \( C_{\text{comp},i} \): This represents the computation capacity vehicular user \( i \) requires for offloading tasks to the edge server.
- \( C_{\text{max},i} \): This is the maximum computation capacity available at the edge server for vehicular user \( i \). It serves as an upper limit for \( C_{\text{comp},i} \).
- \( B_{\text{trans},i} \): This variable denotes the bandwidth vehicular user \( i \) requires for data transmission to and from the edge server.
- \( B_{\text{max},i} \): This represents the maximum available bandwidth at the edge server for vehicular user \( i \). It serves as an upper limit for \( B_{\text{trans},i} \).
- \( E \): This is the total energy consumption for all vehicular users in the network. It includes both computation and transmission.
- \( E_{\text{max}} \): This variable represents the maximum allowable energy budget for the entire network. It serves as an upper limit for \( E \).

**CORA-6G algorithm**

The CORA-6G algorithm aims to optimize computation offloading and resource allocation in 6G vehicular networks. Initially, data on available resources, user requirements, and network conditions are collected, and system utility \( U \), latency \( L \), and energy consumption \( E \) are initialized (Algorithm 1). The algorithm comprises two main components: computation offloading strategy and resource allocation mechanism. For each vehicular user, a convex optimization problem is solved to find the optimal offloading ratio \( P_i \) that maximizes the computation utility \( U_{\text{comp},i} \) while adhering to computation capacity constraints. Subsequently, a two-sided matching algorithm allocates computational and bandwidth resources to maximize \( U \) subject to bandwidth and energy constraints. The algorithm iterates through these steps until \( U \) converges or a maximum number of iterations is reached, ultimately returning the optimized \( P_i, U, L, \) and \( E \).

**Computation offloading strategy**

Computation offloading is a critical component of the CORA-6G algorithm, responsible for determining the optimal offloading ratio \( P_i \) for each vehicular user \( i \) in the network. The offloading ratio represents the proportion of computational tasks that should be offloaded from the vehicular user to the edge server. The computation utility \( U_{\text{comp},i} \) for each user \( i \) is defined as in (4) where \( \alpha \) is the weighting factor for computation utility, \( C_{\text{local},i} \) and

\[
U_{\text{comp},i} = \alpha \cdot C_{\text{local},i}
\]
are the computational capacities for local and offloaded tasks, respectively, and \( T_{\text{local},i} \) and \( T_{\text{comp},i} \) are the time taken for local and offloaded computations, respectively.

\[
U_{\text{comp},i} = \alpha \cdot (T_{\text{local},i} \cdot C_{\text{local},i} + T_{\text{comp},i} \cdot \rho_i \cdot C_{\text{comp},i}) \tag{4}
\]

The objective is to maximize \( U_{\text{comp},i} \) subject to the capacity constraint. In this research, the Lagrangian multiplier method is employed to solve the convex optimization problem of finding \( \rho_i \) for each vehicular user \( i \). This method is particularly useful for optimization problems with inequality constraints. The Lagrangian function \( L \) is formulated by incorporating the objective function and the constraints using Lagrange multipliers \( \lambda \) and \( \mu \) as in

\[
L(\rho, \lambda, \mu) = \alpha \cdot (T_{\text{local},i} \cdot C_{\text{local},i} + T_{\text{comp},i} \cdot \rho_i \cdot C_{\text{comp},i}) - \lambda \cdot (C_{\text{comp},i} - C_{\text{max}}) - \mu \cdot (\rho_i - 1) \tag{5}
\]

Here, \( \lambda \) is the Lagrange multiplier associated with the computational capacity constraint and \( \mu \) is associated with the offloading ratio constraint. Initialized with \( \lambda \) and \( \mu \) set to zero and an initial guess for \( \rho_i \), the algorithm iteratively updates \( \rho_i \) using gradient ascent based on the Lagrangian's gradient. Constraints on computational capacity and offloading ratio are checked and enforced by updating the Lagrange multipliers. The algorithm iterates until the change in \( \rho_i \) falls below a predefined tolerance \( \epsilon \), at which point it returns the optimal \( \rho_i \) that maximizes computation utility while adhering to constraints. Algorithm 2 gives the stepwise computation offloading strategy.

Algorithm 2 Algorithm for Finding Optimal \( \rho_i \) Using Lagrangian Multiplier

1. Input:
2. \( \alpha \): Weighting factor for computation utility
3. \( C_{\text{local},i}, C_{\text{comp},i}, T_{\text{local},i}, T_{\text{comp},i} \): Parameters for computation utility
4. \( C_{\text{max}} \): Maximum computational capacity for user \( i \)
5. Output:
6. \( \rho_i \): Optimal offloading ratio for user \( i \)
7. Initialization:
8. Initialize Lagrange multipliers \( \lambda = 0 \) and \( \mu = 0 \)
9. Set an initial guess for \( \rho_i \), e.g., \( \rho_i = 0.5 \)
10. Set a small positive value for the tolerance \( \epsilon = 10^{-6} \)
11. Step 1: Compute the Gradient of the Lagrangian
12. Compute \( \frac{\partial L}{\partial \rho_i} = \alpha \cdot (T_{\text{local},i} \cdot C_{\text{local},i} + T_{\text{comp},i} \cdot \rho_i \cdot C_{\text{comp},i}) - \lambda \cdot (C_{\text{comp},i} - C_{\text{max}}) - \mu \cdot (\rho_i - 1) \)
13. Step 2: Update \( \rho_i \)
14. Update \( \rho_i = \rho_i + \eta \cdot \frac{\partial L}{\partial \rho_i} \)
15. Step 3: Check and Update Constraints
16. If \( C_{\text{comp},i} > C_{\text{max}} \)
17. Update \( \lambda = \lambda + \delta : (C_{\text{comp},i} - C_{\text{max}}) \)
18. end if
19. If \( \rho_i > 1 \)
20. Set \( \rho_i = 1 \)
21. Update \( \mu \) to a large value
22. end if
23. If \( \rho_i < 0 \)
24. Set \( \rho_i = 0 \)
25. end if
26. Step 4: Check Convergence
27. Compute \( \Delta \rho_i = |\rho_i^{(n)} - \rho_i^{(n-1)}| \)
28. If \( \Delta \rho_i \leq \epsilon \)
29. Return \( \rho_i \)
30. else
31. Update \( \rho_i^{(n+1)} = \rho_i^{(n)} \)
32. Go to Step 1
33. end if

CORA-6G enabled IoT network

In the proposed framework for integrating CORA-6G into an IoT network, IoT devices initially generate data and perform local processing to decide whether to offload computational tasks. If offloading is deemed beneficial, the CORA-6G algorithm is invoked to determine the optimal offloading ratio \( \rho_i \). Tasks are then offloaded to edge servers, which may further offload to cloud servers for computation. Alternatively, tasks are executed locally on the IoT devices if offloading is not chosen. The results from either local execution, edge servers, or cloud servers are subsequently returned to the IoT devices. Figure 1 illustrates the integration of the proposed CORA-6G approach into an IoT framework.

The integration of CORA-6G into the IoT framework aims to optimize network performance through three key objectives. First, it enhances resource allocation by calculating the optimal offloading ratio \( \rho_i \) ensuring efficient task distribution across local devices, edge servers, and cloud servers. Second, it minimizes latency through smart offloading decisions, directing time-sensitive tasks to local or edge servers and offloading computationally intensive tasks to the cloud. Lastly, the algorithm focuses on energy efficiency, optimizing the offloading ratio to reduce energy consumption, especially for battery-operated IoT devices.

This section delineates the experimental framework used to assess the CORA-6G algorithm in IoT-vehicular networks. The focus is on gauging CORA-6G’s efficiency in
optimizing resource allocation and its impact on reducing latency and energy usage.

**Experimental Results and Discussions**

**Experimental Setup**

CORA-6G is evaluated on a 64-bit Windows 10 operating system computer equipped with a distinct hardware configuration, featuring 32.0 GB RAM and an AMD Ryzen 9 5900X CPU with a 4.8 GHz frequency. Various software components, including Python via Anaconda 5.1, are employed to assess the network’s capabilities. To simulate cellular network communication features, the transmission power for each vehicle remains consistent at 100 mW, while offloading tasks’ generation rate ranges from 3 to 6 tasks per minute. Data size and required CPU cycles are randomly generated within the intervals of 30 to 70 Mb and 1200-1600 Megacycles, respectively.

**Performance Evaluation**

The CORA-6G algorithm is evaluated with three objective metrics. The average delay cost (ADC) measures the average time taken for a task to be executed and the result to be returned to the IoT device as in (6). The ADC is crucial for real-time applications where low latency is essential. A lower ADC indicates a more efficient system, which is vital for applications like autonomous driving and emergency response. Here, \( T_{\text{local},i} \) represents the time taken for local computation of the \( i \)-th task on the device itself. It includes the time required for CPU processing, memory access, and other local computational activities for the specific task. \( T_{\text{comp},i} \) denotes the time taken for the computation of the \( i \)-th task when it is offloaded to an external server or edge node. This time includes the data transmission time to and from the server and the actual computation time on the server.

\[
\text{ADC} = \frac{1}{N} \sum_{i=1}^{N} (T_{\text{local},i} + T_{\text{comp},i})
\]  

(6)

The marginal delay sensitivity (MDS) quantifies how sensitive the system performance is to changes in delay as in (7). MDS is significant for understanding the system’s robustness to fluctuations in delay. A lower MDS value implies that the system can better handle variations in latency, which is crucial for maintaining consistent performance.

\[
\text{MDS} = \frac{\Delta \text{ADC}}{\Delta \tau}
\]  

(7)

The average offloading ratio (AOR) measures the proportion of tasks that are offloaded to external computational resources like edge or cloud servers as in (8). AOR provides insights into how effectively the system is utilizing external resources. A higher AOR indicates more offloading, which could lead to better performance but might also increase costs and energy consumption.

\[
\text{AOR} = \frac{1}{N} \sum_{i=1}^{N} \rho_i
\]  

(8)

In the experimental setup, a vehicular IoT network in an urban environment is considered, consisting of varying numbers of IoT devices ranging from 50 to 250. The network also includes 5 edge servers with an average computational power of 10 GHz and 2 cloud servers with an average computational power of 50 GHz. IoT devices’ average local computational power is set at 2 GHz, and the network bandwidth is maintained at 100 Mbps. Table 1 presents the objective metrics for the above setting, which helps to evaluate the performance of the CORA-6G under different network sizes.

This table reveals an intriguing trend where the ADC and MDS decrease while the AOR increases as the number of IoT devices grows. Specifically, ADC drops from 35 to 15 ms, MDS decreases from 0.15 to 0.04, and AOR rises from 0.45 to 0.71 as the network scales from 50 to 250 devices. These trends suggest that the CORA-6G algorithm becomes more efficient and robust with scaling, successfully reducing average delays and becoming less sensitive to latency fluctuations. The rising AOR indicates an increasing reliance on external computational resources, which could signify

<table>
<thead>
<tr>
<th>Number of IoT devices</th>
<th>ADC (ms)</th>
<th>MDS</th>
<th>AOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>15</td>
<td>0.04</td>
<td>0.45</td>
</tr>
<tr>
<td>100</td>
<td>19</td>
<td>0.06</td>
<td>0.51</td>
</tr>
<tr>
<td>150</td>
<td>24</td>
<td>0.11</td>
<td>0.55</td>
</tr>
<tr>
<td>200</td>
<td>32</td>
<td>0.13</td>
<td>0.62</td>
</tr>
<tr>
<td>250</td>
<td>35</td>
<td>0.15</td>
<td>0.71</td>
</tr>
</tbody>
</table>
Dynamic resource management for 6G vehicular networks

both effective resource utilization and a potential area for further optimization. Figure 2 illustrates the above metrics for a clear interpretation of the metrics against increasing network size.

In the context of vehicular networks, channels refer to the communication pathways used for data offloading and vehicle-to-vehicle or vehicle-to-infrastructure communications. The performance of the CORA-6G algorithm is evaluated under multiple channels to understand how increasing the number of available communication pathways influences the data offloading efficiency in vehicular networks. Multiple channels can potentially provide parallel paths for data transmission, reducing congestion and improving overall data throughput.

Figure 3 depicts the performance of CORA-6G for three different values of the number of channels $K = 20, 30, 40$. From the observed results, as the sensitivity to delay increases, the average offloading ratio declines across varying numbers of channels (Figure 4). This suggests that in scenarios where timely data transmission is crucial, there might be a tendency to offload less data as delay sensitivity heightens. A likely reason for this could be to ensure critical data is communicated with minimal delay, even if it means transmitting a smaller volume of data overall.

Systems with more channels, such as $K = 40$, show a notable decrease in offloading ratio as delay sensitivity increases. This suggests that having more channels doesn’t automatically improve offloading in delay-sensitive scenarios. Instead, it indicates a more selective offloading strategy, crucial in applications like autonomous driving requiring real-time decision-making. In dense urban settings, this selectivity helps maintain network stability and efficiency. Thus, while more channels may seem beneficial for data offloading, the trade-off between offloading volume and delay sensitivity highlights the complexity of real-world vehicular networks.

Further, CORA-6G is compared with POETS, the representative work in the context of this research with respect to the number of users. A sharp decline in AOR is evidenced with both approaches as the number of users increases. However, the degradation is highly pronounced with POETS, reinforcing the significance of CORA-6G which maintains a more stable AOR even as the user count escalates. This comparative analysis underscores the robustness of CORA-6G in handling increased network load without a significant drop in offloading efficiency.

The CORA-6G, specifically designed for 6G vehicular networks, enhances network efficiency and performance substantially. One of its most salient features is the incorporation of a convex optimization problem, which is adeptly utilized to determine the optimal offloading ratio.

**Conclusion**

This study explored the paramount significance of efficient computation offloading in 6G vehicular networks, spotlighting the innovative CORA-6G algorithm. The research underscored the algorithm’s ability to adeptly determine the optimal offloading ratio by utilizing a convex optimization problem, thereby maximizing computational...
utility. The findings illuminate how CORA-6G paves the way for enhanced resource allocation and heightened responsiveness, crucial for real-time vehicular applications. Furthermore, the adaptability inherent in the CORA-6G algorithm promises consistent performance, even as network conditions evolve. As vehicular networks advance towards more interconnected and dynamic ecosystems, solutions like CORA-6G will be instrumental in harnessing the full potential of 6G technologies. Future research directions could focus on integrating machine learning techniques to further refine offloading strategies and exploring the scalability of CORA-6G in denser vehicular environments.

References