Privacy- Preserved Oriented Distributed Contextual Online Learning in Mobile Edge Computing

Emran Altamimi*
*Department of Computer Science and Engineering
Qatar University, Doha, Qatar

***Abstract*—This paper presents a cooperative, decentralized multi-UAV framework using distributed online learning to accomplish tasks. The proposed architecture employs cooperative contextual bandits and differential privacy techniques to balance data utility and privacy protection. Asynchronous communication between nodes is facilitated by ZeroMQ's router-dealer pattern, employing TCP as the communication protocol. The framework addresses privacy and security concerns in IoT devices and smart cities, integrating data anonymization and encryption techniques. Performance of these privacy-preserving methods is analyzed and compared to existing literature. Finally, potential areas for future research are discussed, paving the way for more secure, efficient, and privacy-aware UAV swarm systems.**

***Index Terms*—Edge computing, Distributed systems, Differentiated privacy, Multi-armed bandits.**

I. Introduction

Unmanned aerial vehicles (UAVs), commonly referred to as drones, have become increasingly popular in both academic research and industrial applications due to advancements in sensing and computing technology and decreased size and cost. These developments have allowed for the use of UAVs in a variety of applications, including aerial photography, infrastructure inspection, payload transportation, precision agriculture, surveillance, and search and rescue missions. Micro aerial vehicles (MAVs), which are smaller and more agile than traditional UAVs, have expanded the range of applications, but their limited flight time, on-board sensing and computing power, and payload capacity make it difficult for them to perform tasks individually. To overcome these limitations, aerial swarms have been developed in which multiple UAVs work together to accomplish tasks.

The project considers the potential existence of an adversarial environment (e.g., targets may exhibit unpredictability or attempt to deceive the UAV by utilizing new strategies to generate false information in target visitation mission). To tackle this obstacle, we leverage online learning techniques (OL) that empower the UAV swarm to swiftly adapt to abrupt modifications in the environment (e.g., shifts in the probability distribution of targets). This paper introduces a new framework for online learning using multiple cooperative and decentralized UAVs (unmanned aerial vehicles). The framework assumes that each learner receives an instance of data with context information and has the option to either process it using its own processing function or request another learner to process it. The objective of each learner is to learn which processing function will provide the highest total expected reward for that instance. A data stream is a sequence of instances that can only be read once or a few times due to limited computing and storage resources. The data unit can be information obtained from a sensor or camera, while the context information describes the environment or rewards for the learners. Processing functions could include classification functions or transmission strategies depending on the UAV swarm's specific application. Rewards are based on accuracy or good put and energy expended for the selected function or strategy.

We leverage the multi-armed bandits OL technique for each UAV in a cooperative distributed framework to accomplish any task assigned to the UAV (e.g, path planning and target visitation). Multi-armed bandits involve a gambler faced with multiple slot machines, each with different reward distributions. The goal is to maximize total reward by selecting machines to play and determining how often to play them. The challenge is balancing exploration to learn about rewards and exploitation of machines with believed higher rewards. Multi-armed bandit algorithms solve these problems by striking a balance between exploration and exploitation. This framework has applications in various fields, including online advertising, clinical trials, recommendation systems, and reinforcement learning, where sequential decision-making under uncertainty occurs.

The significance of differentiated privacy in distributed systems cannot be overstated, as it offers a structured approach to strike a balance between protecting privacy and enabling data sharing and collaborative processing. In the context of distributed systems, where data is dispersed among multiple nodes or entities, ensuring privacy becomes more complex due to the heightened vulnerability to unauthorized access or data leaks. This paper emphasizes the implementation of differentiated privacy techniques as a means to mitigate the risks associated with data leakage in distributed systems.

In [1], they proposed a novel framework for decentralized online learning involving multiple learners. In this framework, each learner can either select its own action based on the context or request assistance from another learner, with the latter option involving a cost. The learners are modeled as cooperative contextual bandits, aiming to maximize the expected reward from their actions while considering the trade-offs between rewards, information learned, and costs. The paper didn't addressed the problem of security in the proposed distributed system. However, we address the security aspect in distributed online learning using privacy-preserving algorithm

based on securing multi-party computation techniques that allow agents to compute the joint distribution of the contextual features while keeping their individual features private.

In [2], they addressed the privacy and security concerns arising from the large-scale deployment of IoT devices in smart cities, where sensitive data is collected. The proposed approach integrates SDN, data anonymization, and encryption techniques to provide end-to-end privacy and security. This enables efficient and effective data collection and processing, while taking into account the sensitivity of the data and the context in which it is collected. Furthermore, [3] proposed a privacy-preserving intrusion detection framework for SDIoT-Fog environments to ensure the confidentiality and integrity of data collected from IoT devices while addressing the security concerns that arise from the deployment of IoT devices in these environments. The proposed framework uses privacy-preserving techniques to detect anomalies and intrusions in the data while preserving the privacy of the data owner. The framework is integrated with the SDIoT-Fog environment to provide end-to-end security and privacy for the IoT data. Another work considered the privacy preserving introduced in [4] proposed a blockchain-enabled contextual online learning model under local differential privacy for coronary heart disease (CHD) diagnosis in mobile edge computing.To protect privacy, they utilized local differential privacy with a randomized response mechanism and implemented blockchain-enabled information-sharing authentication under multi-party computation.

We address the above aspects as follows:

1) Propose a cooperative, decentralized multi-UAV architecture that relies on distributed online learning (ODL);
2) Leverage the multi-armed bandits algorithm to accomplish the tasks assigned to UAV in the context of DOL.
3) Utilize privacy preserving technique to secure the cooperation between the UAVs in a decentralized manner;
4) Present the performance of the technique in the context of DOL and study the communication efficacy of the system.

II. System Model

In this section, we will discuss the system model for (DOL), performance metrics, and the proposed problem formulation. The DOL architecture, depicted in Figure 1, utilizes UAV swarm technology, which is in high demand due to its ability to complete complex tasks that were previously challenging or impossible for a single UAV to achieve. UAV swarms can collaborate to perform tasks, such as real-time environmental monitoring and tracking, which is often inaccurate when using traditional forecasting methods. Another task is aerial monitoring surveillance, where UAVs can collect intelligence information on detected and identified objects and coordinate their tasks as a fleet to accomplish optimal area surveillance.

In the proposed DOL system, an ad-hoc (peer to peer P2P) network is used to facilitate communication and coordination between the UAVs. In a multi-UAV system with an ad-hoc network, each UAV can act as a node in the network and can communicate directly with other UAVs in its vicinity. This allows for decentralized decision-making and coordination,

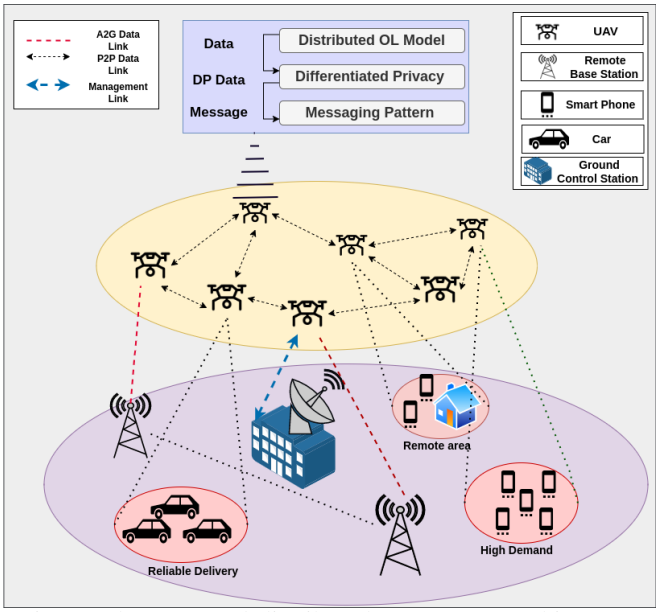

Fig. 1: The proposed distributed OL system architecture.

which can be advantageous in scenarios where a central controller may not be feasible or reliable. The fleet receives the commands from the ground control station, these commands can be initiation of new task, or sends beaming signals to the control station to indicate if the drone is still functioning or otherwise during the pass-through period between a single UAV and the control station. Furthermore, air-to-ground (A2G) links are used to establish communication between the UAVs and a base station for the exchange of information such as sensor data between the UAVs and the base station. This information can be used for various purposes, such as mission planning, situational awareness. Consequently, the model deployed on UAV comprises essential underlying components.

A. Distributed OL

To tackle the challenge of distributed online learning, we utilize the exponential exploration exploitation exp3 algorithm as cooperative contextual bandits. In this approach, a group of collaborative learners has a set of processing functions or ”arms” that can be used to process incoming data. Each learner agrees to follow the exp3 algorithm prescribed by the control station, subject to constraints imposed by the learners themselves. These constraints may include privacy limitations that restrict the amount of information a learner can access about the other learners’ arms. In discrete time, learners receive instances and context information (e.g., context can be the current location of the UAV), which they must process using their own arms or by requesting assistance from other learners, incurring costs in the process. The objective is to maximize the total reward received over a certain period, although learners do not know the expected reward of their own or other learners’ arms. Learners are cooperative because cooperation can lead to mutual benefits, such as increased rewards or learning about arm performance. However, this is a complex problem because learners cannot observe the other

learners' arms or directly estimate the expected rewards of those arms. Moreover, since different learners may receive varying contexts, they may have different learning rates.

B. Differential privacy

Differential privacy (DP) is a privacy model that arose later, in 2006, in the cryptographic community. Unlike k-anonymity, DP was designed to anonymize the outputs of interactive queries to a database. Given $\epsilon \geq 0$, a randomized query function κ (that returns the exact answer to a query function f plus some noise) satisfies ϵ-DP if, for all datasets D1 and D2 that differ in one record and all S $\in$ Range(κ), it holds that :

$$\Pr\left(\kappa\left(D_1\right) \in S\right) \leq \exp(\epsilon) \times \Pr\left(\kappa\left(D_2\right) \in S\right)$$

This equation ensures that the probability of the query function κ applied to dataset D1 being in the set S is at most the exponential of ϵ times the probability of the same function applied to dataset D2 being in the set S. In plain words, the presence or absence of any single record in the database must not be noticeable from the query answers, up to a factor exponential in ϵ. If each record corresponds to a different individual respondent, this means that the individual's information stays confidential. The smaller ϵ, also known as the privacy budget, the higher the protection [5]. DP offers several advantages compared to k-anonymity and the utility-first approach [6], [7]:

- Strong privacy guarantee: DP provides robust privacy protection, largely independent of the attacker's background knowledge. In contrast, k-anonymity requires assumptions about the attacker's knowledge based on the choice of quasi-identifier attributes. While DP is not entirely assumption-free, it still offers a strong guarantee.
- Composability properties: DP has appealing composability properties, which are crucial when dealing with multiple queries or datasets. Sequential composition states that if the outputs of multiple queries are individually protected under ϵ_i-DP, the composed output is protected under the sum of the ϵ_i values. Parallel composition states that if m query outputs computed on m disjoint and independent datasets are protected under ϵ-DP, then the composition of those outputs is still protected under ϵ-DP.

The amount of noise that must be added to satisfy DP depends on the global sensitivity of the query function. Queries with lower sensitivity, such as median or mean, require less noise, while more sensitive functions like maxima or minima need more noise for adequate protection.

In summary, differential privacy is a powerful privacy model that offers strong guarantees and useful composability properties, making it a valuable tool for protecting individual privacy in database queries. Nonetheless, due to its convenient properties and strong privacy guarantee, DP was rapidly adopted by the research community, up to the point that many researchers now consider DP the gold standard in privacy protection.

In the realm of distributed contextual bandits, employing differential privacy may present challenges due to its inherent microdata nature [8]. Nonetheless, the popularity of differential privacy in federated learning settings (which similarly deal with microdata) warrants further investigation into its potential applicability [9]. In the specific case of contextual bandits, Unmanned Aerial Vehicles (UAVs) transmit only selected contexts to other nodes, which suggests that the implementation of differential privacy may not severely degrade the overall data collected by the UAV. In this article, we aim to explore the feasibility of employing differential privacy for safeguarding privacy among UAVs while evaluating its efficacy and impact on the convergence of contextual bandits algorithms within a rigorous framework.

C. Messaging pattern

In order to achieve efficient and secure communication between the nodes in the privacy-preserved oriented distributed contextual online learning framework, we employed ZeroMQ (ZMQ), a high-performance asynchronous messaging library. ZMQ provides a router-dealer pattern, which allows for asynchronous communication between the nodes, ensuring that the system remains responsive even when faced with varying communication delays and processing times. This communication pattern is crucial for maintaining the robustness and adaptability of our mobile edge computing system.

In our implementation, each node is configured to act as both a router and a dealer. This dual functionality allows nodes to send their instance data and request responses and rewards from other nodes (dealer role), as well as to receive, process, and route incoming data and requests from other nodes (router role). This design choice provides a decentralized communication architecture, in which any node can communicate with any other node directly, without relying on a central communication hub. The communication protocol used for this purpose is Transmission Control Protocol (TCP), which provides a reliable, connection-oriented, and error-checked communication channel between nodes.

The communication process begins with a node initiating a connection to another node, followed by the exchange of instance data and request messages. Upon receiving a request, the target node processes it and sends a response and reward back to the requesting node. The contextual multi-armed bandit (CMAB) algorithm then processes the received data to enhance the decision-making process.

In our privacy-preserved oriented distributed contextual online learning framework, we incorporate differential privacy techniques to protect the local data held by each node. By adding a controlled amount of noise to the data before sharing it with other nodes, we ensure individual data points remain private while retaining enough utility for the contextual multi-armed bandit (MAB) algorithm to make accurate decisions. This approach strikes a balance between data utility and the protection of sensitive information, enabling collaborative learning and decision-making without compromising privacy.

In summary, our implementation leverages the router-dealer pattern provided by ZeroMQ to establish asynchronous, decentralized communication between nodes in the mobile edge computing framework. By using TCP as the communication protocol and incorporating security measures such as encryption and message authentication, our system ensures both

efficient and secure data exchange to support the privacy-preserved online learning process.

III. Distributed Multi-Armed Bandits based Solution

In this section, we describe the exp3 algorithm in accordance to [10]. We show the algorithm for single UAV in the network and the behavior of each UAV in the distributed system. We further shows the analysis of the exp3 algorithm.

The UAV can initiate a request and can respond to a coming request from other drones. The request is sent to other neighbouring drones in the network to asses in improving the performance (e.g., in target visitation mission when the environment is extremely cluttered and the drone does not have a full view of the environment, in this case the neighbouring drones might have better views either collectively or individually). A UAV can respond to a coming request and immediate sends the weights to the requester. Algorithm 1 shows the exp3 in the context of DOL.

A. Optimal Learning Rate

The learning rate, denoted as γ, plays a significant role in the algorithm. When the learning rate is set to a large value, the algorithm heavily focuses on the arm with the highest estimated reward, leading to aggressive exploitation. Conversely, for smaller learning rates, the algorithm distributes its attention more evenly, allowing for frequent exploration. It is important to note that as the algorithm concentrates on specific arms, the variance of the estimators for poorly performing arms increases significantly. There are various methods to tune the learning rate, such as adjusting it over time. We simplify the approach by selecting γ based solely on the number of actions, k, and the horizon, T. This restriction implies that the horizon must be predetermined since it influences the chosen value of γ. Equation (1) shows the formula used to calculate the optimal γ in accordance to [10]. k represents the number of actions which can be invoked by the model. T represents the time horizon.

$$\gamma = \sqrt{\frac{\log(k)}{Tk}} \tag{1}$$

B. Regret Analysis

Regret analysis is crucial in bandit problems, where an agent aims to maximize cumulative reward over time. Regret quantifies the shortfall between the agent's total reward and the maximum achievable reward if it always made the best decision, unlike other performance metrics that focus solely on the achieved rewards. The upper bound in Exp3 is important for controlling exploration, balancing exploration and exploitation, enabling regret analysis, and determining the performance trade-offs. It allows for a systematic approach to the exploration-exploitation dilemma and helps achieve better decision-making in our problem. Equation (2) shows the formula used to calculate the upper bound for R in accordance to [10].

$$R = 2\sqrt{Tk\log(k)} \tag{2}$$

Algorithm 1 EXP3 Algorithm

Input: $\gamma \in [0,1]$, T, k.
Set initial $w_i(0)$ to 0, for all $i = \{1, 2, ..., k\}$
for t $\in [0, T]$ **do**
  if Environment is cluttered **then**
    initiate requests to neighbouring drones.
  end if
  Calculate the sampling distribution P_{it}:
  $P_{it} = \frac{\gamma w_{t-1,i}}{\sum_{j=1}^{k} \exp \gamma w_{t-1,i}}$
  Sample A_t from P_t and observe reward X_t
  Update w_{ti}:
  $w_{ti} = w_{t-1,i} + 1 - \frac{\{A_t=i\}(1-X_t)}{P_{ti}}$
  Update the context table with the new w_{ti}
  if request is received **then**
    send w_{ti} to the requester of the current context.
  end if
end for

IV. Performance Evaluation

In this section, we show the performance of the DOL system leveraging the CMAB instance for each UAV in the system. Furthermore, we present the communication efficacy of the system.

A. DOL Performance

In this particular experiment, we utilized a total of three (UAVs), each equipped with its own instance of the "exp3" algorithm. The learning rate was applied to each UAV based on the equation (1), and the regret upper bound was determined using equation (2). The primary objective assigned to each UAV was to accomplish a specific task, where all UAVs within the network were assumed to be engaged in the same task, namely, target visitation. To facilitate task allocation, the coverage area was divided among the UAVs, with each UAV assigned to a designated section. The coverage area was represented by a grid comprising 6x6 cells.

In this experiment, the context considered was the current location of each UAV, which falls under the set of actions denoted as k. Each cell within the grid corresponds to an action within the action space k of the respective UAV. Within this experimental setup, UAV 1 was designated as the requester, while the remaining UAVs served as responders to its requests. The performance of UAV 1 (requester) is depicted in Figure 2.

During the course of the experiment, an exceptionally cluttered environment was observed at time $t = 40k$, leading to a gradual decrease in the weight assigned to cell 2. Consequently, UAV 1 stationed in that area was unable to visually detect the target within that particular cell. To address this situation, UAV 1 initiated a request to the other UAVs, resulting in an increase in the weight associated with cell 2 over time.

Attaining convergence plays a pivotal role in ensuring the stability and effective operation of the system as a whole. The system's stability is evaluated by examining the estimation of regret, as demonstrated in Figure 3 and Figure 4. It

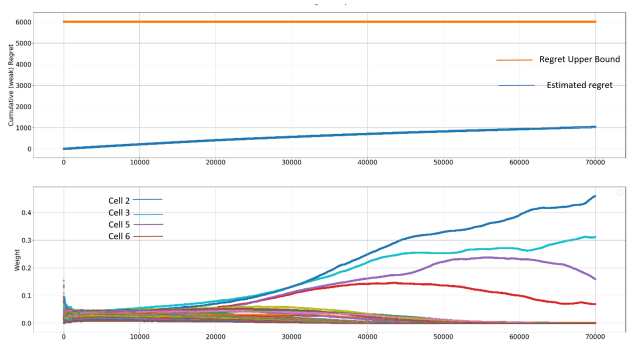

Fig. 2: MAB based target visitation for UAV 1

can be observed that over time, the regret demonstrates a convergent behavior and remains within the prescribed upper bound, indicating that the system is functioning optimally and achieving its desired objectives without exceeding the defined limits.

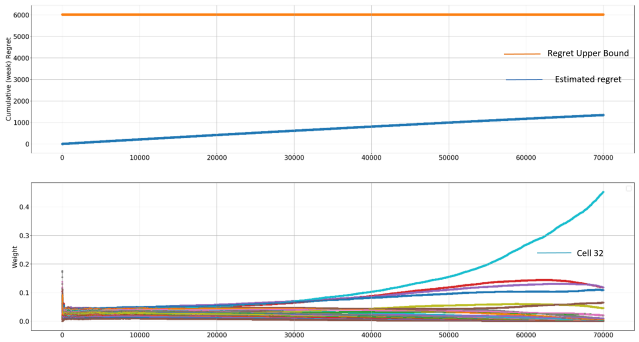

Fig. 3: MAB based target visitation for UAV 2

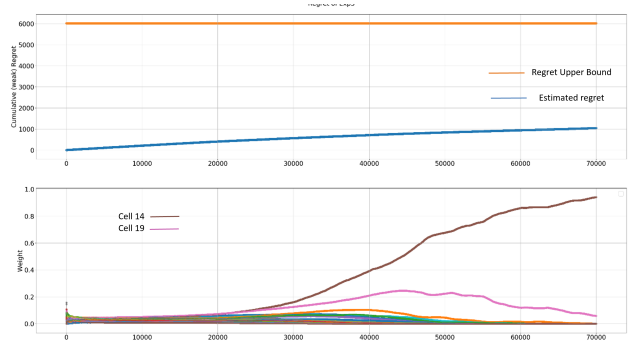

Fig. 4: MAB based target visitation for UAV 3

B. Communication Pattern and Node Implementation

We have used the Router-Dealer pattern for communication between nodes. This pattern is part of the ZeroMQ (ZMQ) library and is particularly suited to our needs. ZMQ is a high-performance asynchronous messaging library, used in distributed or concurrent applications. In our architecture, each node is equipped with a dealer and a router. The dealer sends its data and requests a response. On the other side, the router processes the received data and routes the response back to the requesting dealer. We have implemented this setup using the ZMQ library in Python. We also created a secure node that inherits from the original node class, introducing Transport Layer Security (TLS) for secure communication between nodes. This new class uses the local directory as the certificate authority to verify the identity of the nodes and ensure encrypted communication.

C. Experimental Parameters and Metrics

We have designed our experiments to measure the efficiency and performance of our nodes. Each node keeps track of the following metrics:

- The requests it has sent and the average successful request rate for all nodes.
- The requests it received, normalized by the successful requests sent.
- The responses it has sent, normalized by the requests received.
- The responses it has received, represented as the average successful communication ratio.

These metrics provide a comprehensive view of our network's performance. The experiments allow manipulation of the following parameters:

- Number of nodes
- Size of data to send
- Percentage of nodes that request a response each round
- Percentage of nodes that receive a response each round
- Number of nodes the same request is sent to
- Number of communication rounds
- Communication round time, i.e., the delay between two communication rounds allowing nodes to process the information they have received
- Processing time for the request
- Dealer time out for the response

For secure nodes, the dealer timeout for the response is set to be higher due to the additional time required for secure communication.

We examined the impact of varying the number of nodes, percentage of nodes requesting and receiving responses each round, and communication round time on network performance.

1) Number of Nodes: We tested up to 250 nodes and 5000 messages sent, and found that all messages were received, processed, and routed back successfully for both plain and secure nodes.

2) Nodes Requesting and Receiving Responses: In a scenario where all nodes request but only 10% of nodes receive a response, the successful request sent ratio was 97%. There were no instances where responses failed to be received.

3) Communication Round Time: For a very short round time of 0.5 seconds, the successful sent request and responses received ratios were 0.98 and 0.96, respectively, for the unsecured node and 0.98 and 0.95 for the secure node.

As represented in Figure 6, the plot exhibits the average number of requests, responses sent, and responses received for each communication round. The different parameters set for the simulation, such as the number of nodes, communication round time, and percentage of nodes that request and receive

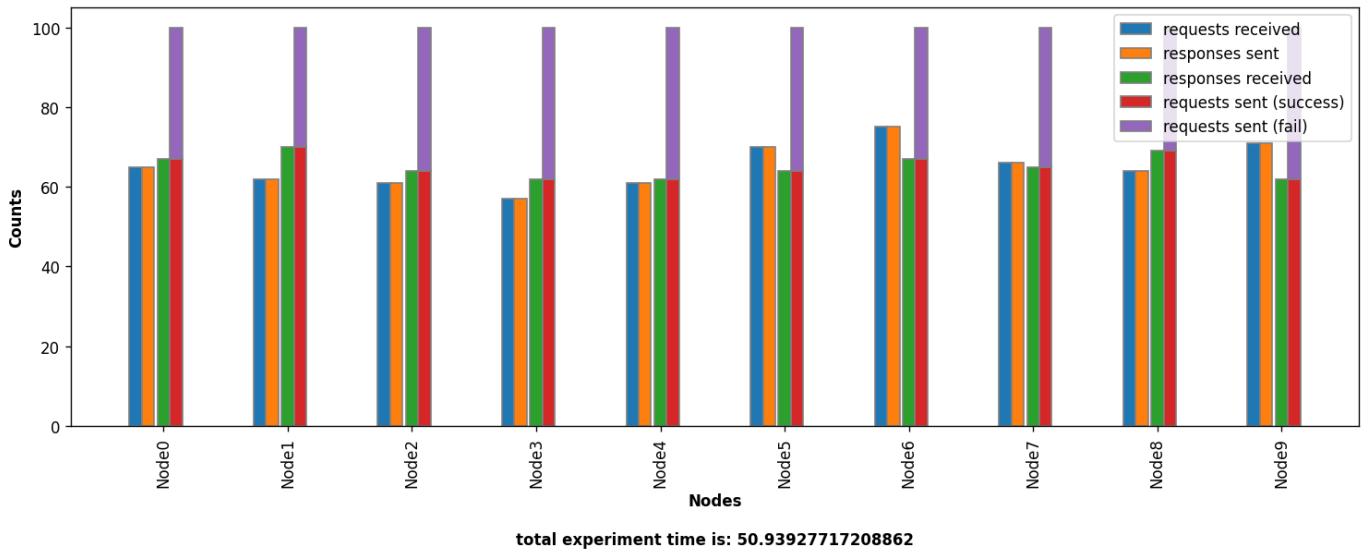

Fig. 5: Comparison of requests received, responses sent, and responses received per node for varying parameters.

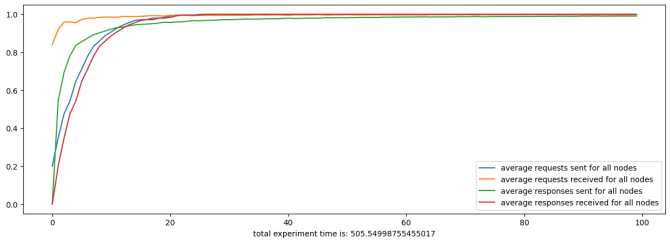

Fig. 6: The average request, responses sent and received for each communication round for varying parameters.

each round, influence these averages. This figure provides an insight into the efficiency and throughput of the network for different configurations, helping us to understand how varying conditions affect the average request-response cycle in the network.

Figure 5 presents a comparative view of the requests received, responses sent, and responses received per node. Again, these metrics are depicted under varying conditions, including the number of nodes, communication round time, and the percentage of nodes that send requests and receive responses each round. This figure underlines the effectiveness of each node in the network, allowing us to observe how they handle requests and responses under different parameters and stress conditions. By analyzing this data, we can fine-tune our system to ensure optimal performance across all nodes, irrespective of the network load and configuration.

V. Conclusion

We introduced a cooperative and decentralized architecture for multiple Unmanned Aerial Vehicles (UAVs), which employed distributed online learning (ODL) based on the multi-armed bandits algorithm. The goal was to enable the UAVs to effectively accomplish their assigned tasks within the framework of DOL. Additionally, we incorporated privacy-preserving techniques to ensure secure cooperation among the UAVs in a decentralized manner. The performance of the proposed technique was evaluated within the context of DOL, with a specific focus on studying the effectiveness of communication in the system. Our observations indicated that within realistic ranges, factors such as processing time, communication rounds, and data size did not significantly impact the overall performance of the system. However, we discovered a strong and direct correlation between the dealer time out and performance, highlighting the crucial role of efficient time management in a distributed system. These findings underscored the importance of effectively managing time constraints for optimizing the performance of the proposed architecture.

References

[1] C. Tekin and M. Van Der Schaar, "Distributed online learning via cooperative contextual bandits," *IEEE transactions on signal processing*, vol. 63, no. 14, pp. 3700–3714, 2015.

[2] M. Gheisari, G. Wang, W. Z. Khan, and C. Fernández-Campusano, "A context-aware privacy-preserving method for iot-based smart city using software defined networking," *Computers & Security*, vol. 87, p. 101470, 2019.

[3] P. Kumar, R. Tripathi, and G. P. Gupta, "P2idf: A privacy-preserving based intrusion detection framework for software defined internet of things-fog (sdiot-fog)," in *Adjunct Proceedings of the 2021 International Conference on Distributed Computing and Networking*, 2021, pp. 37–42.

[4] X. Liu, P. Zhou, T. Qiu, and D. O. Wu, "Blockchain-enabled contextual online learning under local differential privacy for coronary heart disease diagnosis in mobile edge computing," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 8, pp. 2177–2188, 2020.

[5] C. Dwork, "Differential privacy," in *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II 33*. Springer, 2006, pp. 1–12.

[6] C. Clifton and T. Tassa, "On syntactic anonymity and differential privacy," in *2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*. IEEE, 2013, pp. 88–93.

[7] D. Kifer and A. Machanavajjhala, "No free lunch in data privacy," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, 2011, pp. 193–204.

[8] C. Dwork, "A firm foundation for private data analysis," *Communications of the ACM*, vol. 54, no. 1, pp. 86–95, 2011.

[9] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. Quek, and H. V. Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020.

[10] T. Lattimore and C. Szepesvári, *Bandit algorithms*. Cambridge University Press, 2020.