COORP: A High Performance and High Accuracy Wireless Supporting System for Coordinated Robotic Learning

Shengliang Deng, Xiuxian Guan, Zekai Sun, Shixiong Zhao, Tianxiang Shen, Xusheng Chen, Tianyang Duan, Yuexuan Wang, Jia Pan, Yanjun Wu, and Heming Cui*, *Member, IEEE*

Abstract

Coordinated robotic learning (CRL) is a paradigm where outdoor mission-critical multi-robot applications use data parallelism (DP) distributed learning to adapt to new environments, utilizing the distributed computation resources on multiple robots. However, bandwidth-hungry flows initiated by distributed learning contend for the wireless channel intensively, interfering with latency-sensitive flows that are essential for the robots to react timely (e.g. 30ms boundary) to environments. As the widely adopted methodologies of prioritized contention or eliminating contention with global planning cannot support timely transmission and high bandwidth utilization simultaneously, we propose a new methodology: *todo*. We show that reducing contention intensity causes little sacrifice of bandwidth and secures that latency-sensitive flows acquire the channel when contending with bandwidth-hungry flows. On commodity wireless network interface controllers (WNICs), there is another challenge: bandwidth-hungry flows keep occupying the WNIC and block the latency-sensitive flows. We develop a virtual-preemption layer for commodity WNICs to enable latency-sensitive flows to preempt (i.e. get transmitted before) the transmitting bandwidth-hungry flows. Combining these two methods, we propose a novel network supporting system for CRL: *PREemption and DP-Aware contention Reduction* (COORP). The end-to-end evaluation shows that COORP achieves a low violation rate (8.8%) of the reaction time boundary with only a minor slowdown (16.5%) of the training process, guaranteeing high performance and high accuracy of CRL. COORP's code is released on github.com.

*Corresponding author.

I. INTRODUCTION

Coordinated robotic learning enables a group of robots to adapt to unpredictable real-world environments and is increasingly important to mission-critical multi-robot applications in field, including rescue [1], navigation [2], and surveillance [3], [4]. Despite each robot's local functionality (e.g., collision detection), we summarize that a typical coordinated robotic learning workflow [1], [2], [4]–[9] comprises two unique global coordinations: a *multi-view perception-control coordination* (MPC) that continuously exchanges environmental information among robots and makes coordinated decisions, and a *distributed reinforcement learning coordination* (DRLC) that exploits the distributed computation power of all robots to adapt to new environment.

The MPC of a robot group loops over *perception exchange*, *inference*, and *control dissemination*, as shown in Figure 1. In each iteration, during *perception exchange*, a leader robot collects and combines perceptions (e.g., positions or images) captured by the sensors (e.g., cameras) of all the other robots, via a wireless network for multi-view awareness of the environment. During *inference*, the leader robot feeds the combined perception to a control model which takes environmental information as input and produces cooperative control messages (e.g., velocities and angles) as output. After that, the leader robot *disseminates* the control messages to each robot for the next move.

The DRLC is usually spawned when a robot group enters a new environment [5]. As the hardware resources per robot are limited, data parallelism (DP, see Section II-A) is usually adopted to exploit the distributed computation power of all the robots. A data parallel DRLC consists of *data dissemination*, *parallel training*, and *parameter synchronization*. Note that, in each iteration of the MPC loop, the leader robot collects a composed data of a combined perception, control messages that contain actions, and a corresponding reward that reflects the effect of the actions. When a batch of data is accumulated, the leader robot partitions the data batch and *disseminates* each robot a partition of the data. Then each robot trains the control model *in parallel*. The leader robot gathers the parameter updates computed by each robot to *synchronize* the control model. The learning process repeats until the robot group adapts to the new environment. In this paper, we focus on coordinated robotic learning applications with one MPC and one DRLC.

We identify two stringent requirements for mission-critical CRL applications. First, the MPC loop requires *fast reaction* (**R1**): each iteration of MPC loop should finish within a tight reaction

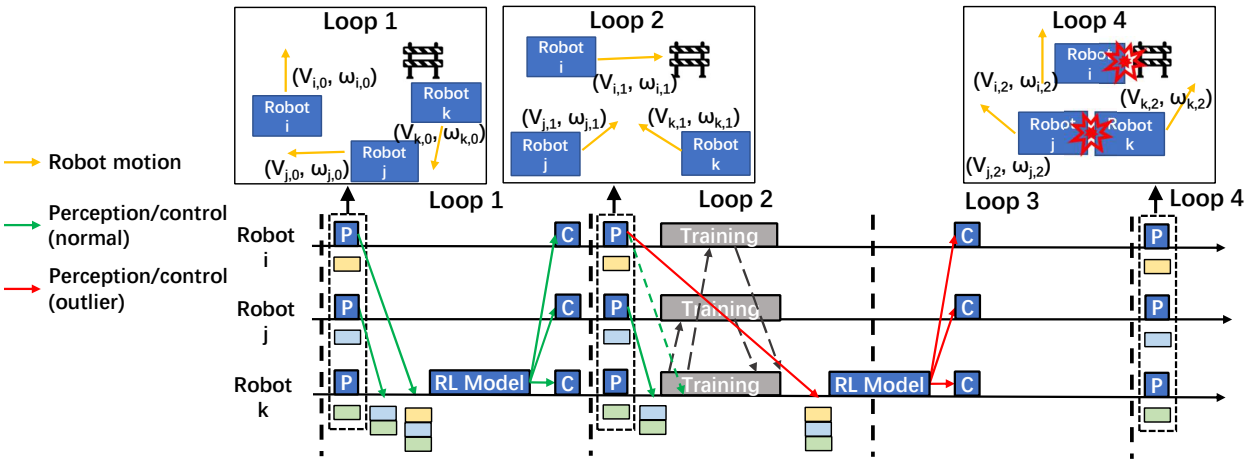
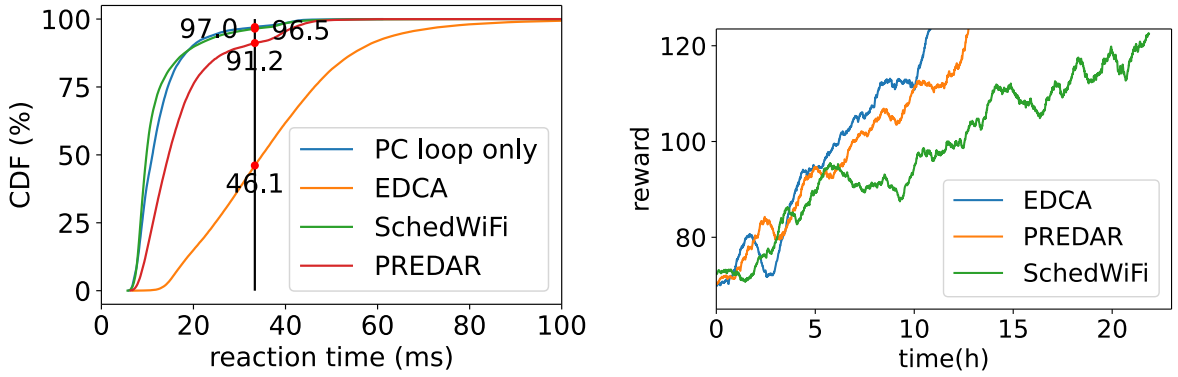


Fig. 1: Workflow of coordinated robotic learning and why fast reaction is important. 'P' stands for 'perception'. 'C' stands for 'control'. Bandwidth-hungry flows for training block the latency-sensitive flow (red line) and causes violation of the reaction time boundary (loop 2), leading to lag and collision (loop 4).

time boundary (e.g. 33ms [10]–[12]). Fast reaction avoids severe flaws like robot collisions demonstrated in Figure 1. Second, the DRLC requires *fast model convergence* (**R2**): the control model should adapt to a new environment (i.e., converge) as fast as possible.

However, the two coordinations (MPC and DRLC) generate network traffic flows with different characteristics, making it cumbersome for the underlying network supporting systems to meet these two stringent requirements simultaneously. In terms of traffic volume, the messages of MPC (i.e., perception and control messages) are often in small scale (e.g., several KB); while the messages of DRLC (i.e., data dissemination and parameter updates) are often in much larger scale (e.g., tens to hundreds of MB). Still, in terms of transmission quality, the messages of MPC are generated periodically (e.g., 30Hz) and the transmission of each message is latency-sensitive (LS): a latency that violates the reaction time boundary may cause flaws (i.e., violate **R1**); while the messages of DRLC are generated less frequently but are bandwidth-hungry: a limited available bandwidth will slow down the speed that a distributed learned control model adapts to a new environment (i.e., violate **R2**).

Unfortunately, no existing wireless network supporting system meets **R1** and **R2** simultaneously when serving these two aforementioned traffic flows in coordinated robotic learning. Existing systems can be classified into two categories. The first category is *prioritized contention*



(a) Reaction time.

(b) Convergence over time.

Fig. 2: Comparison of different systems. In 'MPC loop only', only the MPC loop process was running. COORP achieved a low percentage of violation of the reaction time boundary (vertical black line) with only a minor slowdown of the training process.

systems [13]–[17]: each wireless network interface card (WNIC) on each robot occupies the channel when the upper layer passes messages for it to transmit; WNICs that have LS messages to transmit are given higher chances to occupy the channel before other WNICs. Prioritized contention enables robots to communicate without central control and achieves high bandwidth utilization. However, it fails to achieve **R1** in CRL because multiple WNICs with BH messages to transmit simultaneously contend for the channel, lowering the chance for WNICs with LS messages to timely acquire the channel and causing latency increase (Section II-B).

The second category is global planning systems [18]–[20]: a global scheduler divides the time into time windows and arranges the transmission of LS messages and BH messages into exclusive time windows. Global planning provides low latency in cases that a limited available bandwidth does not affect the service quality of the application.

However, in dividing and assigning the time window, global planning has to consider uncertainties from two potential sources: first, the packet transmission time is not stable in wireless networks; second, dividing a time axis across multiple devices requires a global clock but faces clock skewness [18]. In the presence of these uncertainties, global planning faces a paradox to meet both **R1** and **R2** in CRL: on the one hand, as the scheduler cannot collect the uncertainty on each robot in real-time, it has to reserve a large enough time window for each LS message transmission, taking up too much of the bandwidth, violating **R2**; on the other hand, reducing

the size of each time window to meet **R2** would make LS transmissions more likely to miss their own windows, violating **R1**.

For instance, we ran a route planning coordinated robotic learning application on five robots with SchedWiFi [20], the most relevant global planning system that is designed for low latency of control messages in factory automation where high bandwidth utilization is usually not required. Despite a low percentage (3.5%) of violation of the reaction time boundary (Figure 2a), SchedWiFi suffered a 85.3% slowdown of the training process (Figure 2b) due to a 41% reduction of available bandwidth (Table II).

Overall, a system meeting R1 and R2 simultaneously for coordinated robotic learning is highly desired but missing. In this paper, we build COORP, the first network supporting system for cooperating the two types of coordinations and meeting both **R1** and **R2** for coordinated robotic learning. Lying in the core of COORP is our newly proposed abstraction: *coordinated preemption*. Coordinated preemption should ensure that whenever an LS message from any of the robots is to be transmitted: locally, the WNIC on the robot should immediately serve the LS message for transmission (*local preemption*); globally, the WNIC which serves the LS message should immediately acquire the wireless channel shared by all the WNICs of the robot group (*global preemption*). COORP meets **R1** by enforcing the coordinated preemption. Still, COORP's coordinated preemption maintains a reasonably high available bandwidth to meet **R2**.

Enforcing *global preemption* is challenging: as multiple WNICs are contending the same wireless channel, when an LS message is to be transmitted from a robot's WNIC, the wireless channel might already be contended by multiple WNICs that are serving a high volume traffic; this lowers the chance that the WNIC with an LS message acquires the channel even with EDCA [17], a technique since WiFi 4 that makes a contention incline to a WNIC assigned with higher priority; thus, the LS message is blocked (violate **R1**). To ensure global preemption for coordinated robotic learning, we leverage a key observation that, in the *data dissemination* and *parameter synchronization* of the DRLC, the critical path is dominated by the transmission time of all the BH messages via a single wireless channel. Leveraging this observation, COORP minimizes the contention by only allowing one BH message to be transmitted at any time. By doing so, at any time, there is at most one WNIC with BH messages contending for the channel, and when any LS message is to be transmitted, it can easily win the contention with the highest chance (meet **R1**). Moreover, COORP lets BH messages saturate the wireless channel in a round-robin manner which ensures that the overall bandwidth utilization is high (meet **R2**).

Enforcing *local preemption* on commodity wireless network interface controllers (WNICs) is confronted with a unique challenge that, in modern WiFi protocol [21], when a large number of packets (segmented from a message) are being served, newly arrived packets have to wait until the WNIC finishes the transmission of a number of previous packets due to frame aggregation [22], a key technique for modern WiFi to provide high bandwidth. Our key observation for this challenge is that, in the MPC of coordinated robotic learning, the LS messages are often generated at a constant frequency which is predictable by COORP. Leveraging this observation, we propose a novel virtual-preemption layer that predicts when an LS message will arrive and limits the number of packets of BH messages the OS passes to the WNIC, such that when an LS message arrives, previous packets buffered by the WNIC have been finished. By doing so, an LS message can be served immediately by the WNIC whenever it is generated (meet **R1**). COORP dynamically estimates the maximum number of packets that could be transmitted before the arrival of the next LS message to meet **G2**. The virtual-preemption layer leverages common operations of the WNIC driver, thus it is portable to different commodity WNICs.

We implemented COORP with around 1000 lines of code in Linux 5.4.84 and ROS2 [23]. We compared COORP with two representative methods, EDCA [17] and SchedWiFi [20], on a typical route planning coordinated robotic learning application [2] using commodity WNICs. Evaluation shows that:

- COORP achieves fast reaction. COORP realizes a low percentage of violation (8.8%) of reaction time boundary, comparable to the lowest possible case (3.0%) with the same hardware.
- COORP achieves fast model convergence. COORP only requires 16.5%~xxx more time than the fastest baseline (EDCA) to obtain a comparable control model.

Our major contribution is COORP, the first network supporting system that fulfills both *fast reaction* (**R1**) and *fast model convergence* (**R2**) for coordinated robotic learning. COORP enables coordinated robotic learning applications to adapt to new environments with minimal time while avoiding lag and collisions among robots. With COORP, more coordinated robotic learning applications can be developed for complex real-world tasks that require close coordination among robots.

In the rest of this paper: Section II discusses related work; Section III presents the system model and overview of COORP; Section IV describes the design of COORP; Section V describes the implementation; Section VI shows our evaluation; Section VII concludes.

II. Background and Motivation

A. Coordinated Robotic Learning

Multi-view Perception-Control Coordination. Traditional robot control algorithms mainly leverage local perceptions collected with sensors on each robot [24], [25]. While these algorithms perform well in single-robot tasks, when multiple robots work together for a common goal, they are unable to leverage perceptions of other robots (e.g., image of an obstacle from different angles) and may lead to suboptimal decisions.

With more and more research efforts in multi-robot applications, a multi-view paradigm emerged: multiple robots in a group combine their perceptions of the environment and make coordinated decisions based on the combined perceptions [2], [4], [7]–[9] as described in Section I. For example, in [2], the authors proposed a multi-robot navigation algorithm that combines perceptions of all the robots to get the state of all the obstacles, and generates coordinated actions. Note that, without combining perceptions of all the robots, each robot would only know the position of its nearby obstacles and take suboptimal actions. In real-world deployments, such multi-view algorithms may coexist with traditional single-view algorithms for the flexibility of the entire robot group [24].

Distributed Reinforcement Learning Coordination. Deep reinforcement learning trains a deep neural network control model by interacting with the environment, accumulating experiences, and adjusting the deep neural network with the experiences, and is especially effective for robot control involving interaction with complex environments [26], [27]. Despite the flexibility and learning capability of deep reinforcement learning, even a well-trained control model requires fine-tuning in new environments [28], [29], which is computation intensive and memory consuming.

Data parallelism is widely adopted in training deep neural networks to alleviate the computation and memory requirements on each single worker. In data parallelism, the training data is partitioned and scattered to multiple workers. These workers train the same control model using their own partition of the data in parallel and synchronize the model parameters with others. There are different paradigms for synchronizing the parameters across workers [30], namely Bulk Synchronous Parallel (BSP), Asynchronous Parallel (ASP), and Stale Synchronous Parallel (SSP). These paradigms differ in when to synchronize the parameters, but they all transmit training data and parameter updates over the network and incur large volume traffic.

B. Related Work

The WiFi protocol uses distributed contention to decide which WNIC transmits its packets over the shared wireless channel: each time a WNIC wants to initiate a transmission, it first sets a counter to a randomly chose number between 0 and a pre-defined parameter CW; then it decrements the counter when the channel is not used by others until zero before it actually initiate the transmission. This requires little coordination among WNICs and is easy to make different wireless networks coexist, thus it is widely adopted. Despite the convenience, the distributed contention is unable to provide low transmission latency for latency-sensitive traffic. Existing research efforts can be classified into two categories: *prioritized contention* and *global planning*.

Prioritized Contention. This category improves the contention process such that an LS message has a higher chance to win in the contention. For example, EDCA [17], a part of modern WiFi protocol, sets up four traffic categories, namely `AC_VI`, `AC_VO`, `AC_BE`, and `AC_BK`, and assigns different initial CWs for these traffic categories, such that higher priority traffic can gain earlier access to the shared wireless channel. [14] dynamically limits the queue length on each host such that LS packets can bypass most of the packets waiting to be transmitted, but the LS packets still have to wait for packets that are already passed to the WNIC and also contend with transmission from other WNICs. While these systems are effective in daily use cases, they are not suitable for coordinated robotic learning, since multiple robots have BH flows to transmit simultaneously, each saturating the limited bandwidth, causing intensive contention and significantly lowers the chance for LS packets to timely access the channel.

Global Planning. This category leverages a global scheduler to divide the time into time windows and arrange the transmission of LS messages and BH messages into exclusive time windows. For example, TDMA-based approaches [19], [31]–[33] proactively assign time windows to different wireless stations. While the contention is fully avoided, they lack flexibility to accommodate dynamically generated traffic like those sent by TCP. Also, they require all the wireless stations to have a global synchronized clock, thus incur additional synchronization overhead and are vulnerable to clock skewness [18]. [18] alleviates the need of a global clock by polling each station to transmit. It achieves high efficiency and fairness across wireless stations. However, for a low transmission latency, the global scheduler still needs to know when an LS message is generated, requiring either additional coordination or some bootstrap.

C. Motivation

In this work, we first propose a new abstraction named *coordinated preemption* consisting of *local preemption* and *global preemption*, then realize it specifically for coordinated robotic learning with COORP.

Global Preemption. Global preemption ensures that, whenever an LS message is to be transmitted by a WNIC, the WNIC immediately acquires the wireless channel shared by all the WNICs of the robot group. We notice that the design of EDCA is effective when a small number of BH flows contend for the shared wireless channel. Meanwhile, even a single BH flow is able to saturate the available bandwidth due to the relatively constant network capacity and the large traffic volume of BH flows. Further, in the data dissemination and parameter synchronization of the DRLC, the critical path is dominated by the transmission time of all the BH messages for synchronization via a single wireless channel. Thus, it is viable to reduce the contention by only allowing a single BH flow to be transmitted at any time, and let them saturate the available bandwidth in a round-robin manner.

Local Preemption. Local preemption ensures that, on each robot, whenever an LS message is to be transmitted, the WNIC on the robot immediately serves the LS message for transmission. Although this was well-studied as a typical problem of the OS network stack [34]–[41], a pre-condition for existing work to be effective no longer holds on commodity WNICs: *the OS network stack must be the last component where the LS message can be blocked*. Modern WNICs work asynchronously with the OS to achieve high throughput with frame aggregation. Packets delivered to the WNIC by the OS accumulate in the WNIC's internal buffer and keep the WNIC busy. Newly delivered packets have to wait until the ongoing transmission of previous packets is finished, making them blocked in the WNIC, as conceptually depicted in Figure 3. We term this as *local congestion*. Local congestion exists even if the WNICs have multiple hardware queues for different access categories defined by EDCA, because the hardware components for transmission are saturated by the BH packets when LS packets are delivered to the WNIC. Further, local congestion gets amplified when there are BH flows from other robots.

Figure 4 demonstrates local congestion and its amplification. The experiment was carried out with the same testbed as described in Section VI, and FQ-CoDel [34], a queue management algorithm that optimizes for latency-sensitive sparse traffic, was adopted by default such that the LS messages were not blocked in the OS network stack. The latency-sensitive messages

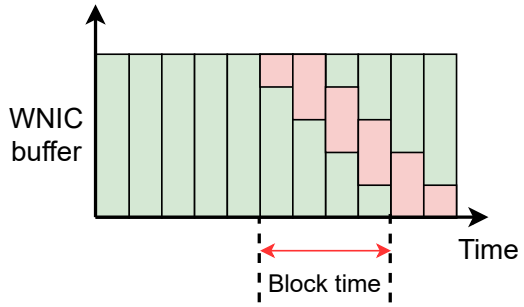


Fig. 3: Conceptual depiction of the congestion in the WNIC. Packets of BH flows delivered by the OS accumulate in the WNIC buffer and keep the WNIC busy transmitting. Packets of LS flows have to wait until transmission of previous packets is finished.

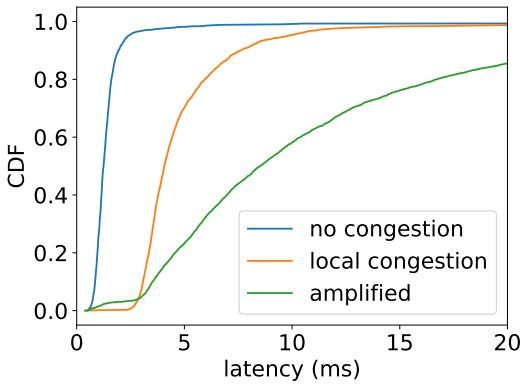


Fig. 4: Local congestion and its amplification. In 'amplified', local congestion is amplified by a BH flow from another robot. Both local congestion and its amplification cause a significant latency increase.

were configured to transmit through `AC_VO`, and the bandwidth-hungry flows were configured to transmit through `AC_BE`.

In COORP, local preemption is achieved with the virtual-preemption layer which enables LS flows to virtually preempt local BH flows to avoid the increase of latency. This layer leverages common driver operations (see Section V) and requires software-only modifications to the OS, thus it supports different commodity WNICs.

III. SYSTEM OVERVIEW

A. System Model

In the deployment model of COORP, two coordinations, i.e., MPC and DRLC as described in Section II, are deployed in a robot group: the MPC controls the robots to closely coordinate for a common task, and the DRLC adapts the robot group to new environments. Each robot is equipped with sensors like cameras and LiDARs that generate perceptions. Among all the robots in the group, a leader robot gathers the perceptions of all the robots and holds a deep neural network as the control model to produce cooperative control messages for each robot. The other robots (worker robots) are within one-hop distance from the leader robot so that the communication between the worker robots and the leader robot does not need any relay. All

robots communicate via a wireless network in a shared channel. RTS/CTS [42] is enabled in the wireless network to avoid interference caused by hidden terminal problems [43]: multiple worker robots transmit packets simultaneously and interfere with each other.

In the MPC, for the multi-view coordination, the perceptions are generated from the sensors on each robot periodically. Time is divided according to the generation period (e.g., 33ms) of perceptions. Each period is expected to contain exactly one MPC loop. A reaction time boundary no longer than the generation period is configured, and MPC loops are expected to finish within the reaction time boundary. An MPC loop initiated in a period can lag and finish in the next period or later as shown in loop 3 in Figure 1, and put off subsequent MPC loops.

In the DRLC, the leader robot collects a training data item (i.e., perceptions, control messages, and rewards) per MPC loop. The rewards are computed according to the subsequent perceptions. When a certain number of training data items are collected, the leader robot disseminates partitions of training data among robots and continues the concurrent MPC loop process without collecting training data, because these training data items are generated under the old control model. When parameter synchronization finishes and the control model is updated, the leader robot resumes collecting new training data under the updated control model.

B. Overview of COORP

COORP fulfills the two requirements of coordinated robotic learning (i.e., *fast reaction* and *fast model convergence*) by enforcing our newly proposed abstraction, *coordinated preemption*, while maintaining a high available bandwidth. Coordinated preemption ensures that whenever an LS message from any of the robots is to be transmitted: locally, the WNIC on the robot immediately serves the LS message for transmission (local preemption); globally, the WNIC which serves the LS message immediately acquires the wireless channel shared by all the WNICs of the robot group (global preemption).

Figure 5 shows the architecture of COORP. On each robot, a BH controller and a preemption controller reside in the user space, which intercepts the generation of bandwidth-hungry flows and latency-sensitive flows respectively. A global controller resides in the user space of the leader robot. A virtual-preemption layer wraps the WNIC driver in the kernel space on each robot, which enables LS flows to preempt BH flows on the WNIC.

a) ***Enforcing Global Preemption****:* As discussed in Section II, the concurrently generated BH flows from different robots in the DRLC contend for the channel intensively and increase the

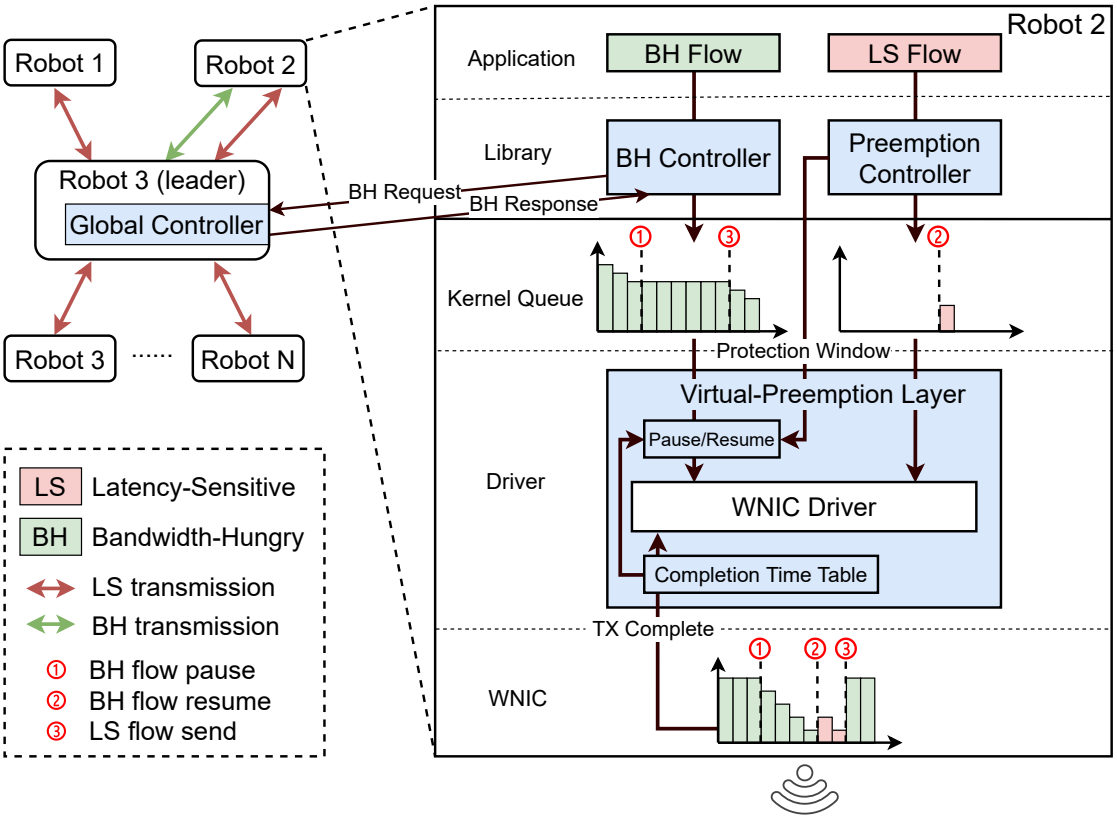


Fig. 5: The architecture of COORP. COORP's components are shaded in blue. ① The virtual-preemption layer pauses the kernel queue of the BH flow to avoid passing too many packets to the WNIC; the WNIC still transmits the packets it has buffered. ② The application sends an LS message; the WNIC serves the LS message immediately because it has finished previous packets passed by the OS. ③ The virtual-preemption layer resumes the kernel queue of the BH flow because the transmission of the LS message is finished. Meanwhile, the global controller and BH controllers ensure there is at most one BH flow being transmitted at any time.

latency of LS flows, violating **R1**. We observe that reducing contention intensity is a promising way to enforce **R1** and **R2** simultaneously among flows from different robots (verified in Section II). A strawman approach to reduce contention intensity is to leverage a time division multiple access (TDMA) method: dividing time into time slices and exclusively assigning time slices to each BH flow. However, the TDMA method is unsuitable for coordinated robotic learning because different BH flows in the learning process can start and finish transmission at different times due to the variation of training speed and transmission throughput. Once certain BH

flows finish transmitting, the future time slices and bandwidth assigned to them will be wasted (violating **R2**).

Instead, COORP leverage a request-based semi-TDMA method that requires minimal coordination. If a robot attempts to transmit BH flows, its BH controller in the user space will intercept the transmission and request the global controller for permission to transmit. If there is no other BH flow transmitting, the global controller will grant the transmission permission with a response immediately; otherwise, the response will be postponed until the transmitting BH flow finishes or pauses. This process only needs a request message and a response message, which occupies little of the bandwidth compared to the large size of parameter updates, guaranteeing **R2**. Moreover, reducing the number of concurrently transmitting BH flows would not lower bandwidth utilization because a BH flow will take up as much bandwidth as possible.

*b) **Enforcing Local Preemption**:* Due to the closed-source design of commodity WNICs, the transmission on the WNIC cannot be manipulated directly. To enforce local preemption, we design and implement a software-only virtual-preemption layer surrounding the WNIC driver between the WNIC and upper kernel layers to virtually enforce preemption for LS flows. The virtual-preemption layer relies on a preemption controller to anticipate the arrival time of LS packets (i.e., the time when an LS flow would send a message). The virtual-preemption layer records the time that BH packets take to be transmitted by the WNIC (transmission completion time) and further estimates the future transmission completion time of new BH packets to be fed to the WNIC. If the transmission completion time of a BH packet would overlap with the arrival time of LS packets, the virtual-preemption layer will stop feeding the BH packet to the WNIC until the LS packets arrive. In this way, preemption on the WNIC is virtually achieved and LS packets would not be blocked by the BH packets on the WNIC, minimizing the transmission latency (**R1**). Moreover, the transmission completion time is collected in real-time from the driver which is close to the WNIC. Thus the time to stop BH flows is accurate (benefitting **R1**) and minimal (benefitting **R2**).

IV. DETAILED DESIGN

A. Global Controller and BH Controller

In COORP, a request-based semi-TDMA method as shown in Algorithm 1 is integrated with EDCA enabled, so that contention intensity among BH flows is reduced and LS flows can acquire the channel with high probability when contending with BH flows (verified in Section II). The

global controller on the leader robot maintains a queue Q of robots that requested to transmit BH flows and a set S of robots that are transmitting BH flows concurrently.

First, when a worker robot needs to transmit a BH flow, its BH controller would send a BH request to the global controller on the leader robot and wait for a response (line 13) before transmission. Second, the global controller would allow the requested transmission by sending a BH response when the number of concurrently transmitting BH flows in S has not reached a limit (e.g., 2 BH flows), so that EDCA can effectively prioritize the transmission of flows according to their latency requirements (**R1**). If the limit is reached, the global controller would add the request to Q and postpone the response until the number of concurrently transmitting BH flows decreases (line 8). Third, on response, the robot would keep transmitting BH flows for a pre-configured time slice (e.g., 5s), or if the robot finishes transmitting BH flows within the pre-configured time, it would notify the global controller to recycle the unused time for other BH flows.

The set S is maintained by the global controller in the following ways: if the global controller permits a robot to transmit BH flows, the time is recorded and the robot is added to S; if a robot has been transmitting for the pre-configured time according to the permission time records, or a robot notifies that it finishes transmitting BH flows (release), the robot is removed from S.

B. Preemption Controller

Each time an LS flow sends LS packets, the preemption controller monitors the generation time of LS packets at the user space and records it as a sample. Suppose an LS flow started from time q and its period is p, the ideal arrival time (i.e., when it is generated by the application) of the kth ($k = 0, 1, 2, ...$) message can be modelled as $\hat{T}_k = pk + q$. However, due to random fluctuations, there is a difference between the actual time T_k and the ideal time $\hat{T}_k$. As shown in Figure 6, we estimated that the fluctuation roughly follows a normal distribution $N(0, \sigma^2)$: $T_k \sim N(pk + q, \sigma^2), k = 0, 1, 2...$ The parameters p, q and σ constitute a time model $\{p, q, \sigma\}$ of an LS flow.

With enough samples of an LS flow, the preemption controller runs linear regression on these samples to estimate the corresponding time model of an LS flow. Since the inferred parameters may become inaccurate over time, the preemption controller recomputes the model in two conditions: 1) the difference between the predicted time and the actual time of a message exceeds 2σ; 2) a statically configured time interval has passed since last recomputing. With

Algorithm 1: Request-Based Semi-TDMA

```
1  Function schedule(): ;                                  /* global controller */
2
       Data: Q: queue of robots requesting for transmission; S: set of robots transmitting
3      upon get request from robot r do
4          Q.put(r);
5      upon robot r in S for more than TIMESLICE
6           || get release from robot r do
7          S.remove_if_exist(r);
8      upon r ← Q.get() do
9          wait until |S| < LIMIT;
10         send permit to r;
11         S.add(r);
12 Function send(data): ;                                  /* BH controller */
13
14     while data not all sent do
15         request_wait();
16         transmit data for no more than TIMESLICE;
17         release();
18     end
```

the time model $\{p, q, \sigma\}$, the future arrival time T_k of an LS flow can be easily inferred as $[\hat{T}_k - \Delta, \hat{T}_k + \Delta]$ according to normal distribution under a given accuracy requirement denoted as C. We set C to 95% for which the Δ can be approximated with 2σ.

We place the preemption controller at the userspace because a robot can generate multiple LS flows and their time models may vary. The preemption controller has to infer the time models of different LS flows respectively, which is more convenient at the userspace where the flows are separated.

The design of the preemption controller has to cope with two challenges: first, the future arrival time of each LS flow may overlap with each other; second, the future arrival time should be reported to the virtual-preemption layer timely, otherwise there will not be enough time for the

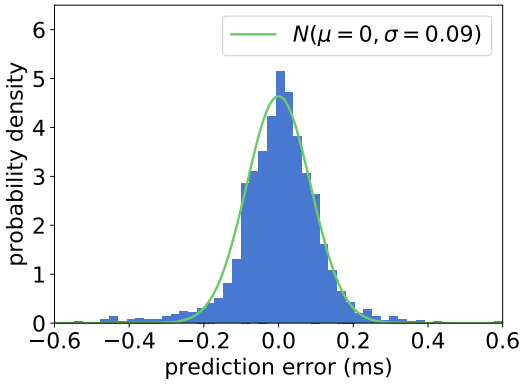


Fig. 6: The prediction error roughly follows normal distribution.

WNIC to finish its ongoing transmission, causing local congestion. To tackle these challenges, we use a protection window (PW) that covers the future arrival time of multiple LS flows. Suppose the current future arrival time of all LS flows constitutes a set T and the earliest one is $t = [t_1, t_2]$, we set the PW to t, remove t from T and look up the new earliest one t' in T. If $t' = [t'_1, t'_2]$ overlaps with PW (i.e., $t_1 \leq t'_1 \leq t_2$), t' is combined into PW (i.e., PW= $[t_1, max(t_2, t'_2)]$) and removed from T. The process repeats until a new t' does not overlap with PW, so that PW starting from t covers all future arrival time that needs to be protected as a whole. Upon the ending of a PW, the preemption controller will update the future arrival time of each time model of each LS flow and calculate a new PW; the preemption controller will notify the virtual-preemption layer to resume the paused BH flows and update the new PW simultaneously. In this way, the virtual-preemption layer can always control the feeding of BH packets under the latest PW.

C. Virtual-preemption layer

The functions of a driver can be concluded into three major operations: dequeuing packets from upper kernel queues; enqueuing packets to the WNIC; receiving the transmission completion callback of a packet. The virtual-preemption layer wraps the driver by integrating with all these three operations as shown in Algorithm 2.

The virtual-preemption layer maintains a completion time table (CTT) that is used to predict the required time for transmitting specific numbers of packets on the WNIC. The CTT contains rows with form (n, t) which means that it takes time t to complete the transmission of $n + 1$ BH packets on the WNIC. To maintain the CTT, the virtual-preemption layer intercepts into

both the enqueue and the transmission completion callback of the WNIC driver (line 11 and line 16). Each time the driver is enqueueing a BH packet into the WNIC's BH queue, the enqueue time t_{enq} of the packet is recorded, along with the current number of packets n in the WNIC's BH queue. Suppose the packet transmission is completed at time t_{comp}, we learn that the BH queue needs $(t_{comp} - t_{enq})$ time to complete the transmission of a packet together with n packets previously queued in the WNIC queue (line 20). As the completion time varies due to instability and contention in the wireless channel, we keep a number of the most recent records for each n (ranging from 0 to the maximum queue depth of the WNIC), and empirically take the maximum as the estimated transmission time.

Given a PW denoted as $[t_1, t_2]$ and current time T, if $t_1 \geq T$, the virtual-preemption layer would protect the LS flows that are supposed to arrive within the PW by intercepting the dequeue operation of the driver (line 1). When the driver attempts to dequeue a BH packet from the upper kernel queues, the virtual-preemption layer would query the number of packets currently queued on the WNIC n_{pkt}, and query the corresponding transmission completion time t_n from the CTT (line 2-3). If $t_n > t_1 - T$ or $t_2 > T > t_1$, the transmission of this BH packet is not likely to be finished before the arrival of the LS flows and would block the transmission of LS flows (line 5). Thus the virtual-preemption layer would abort the dequeuing operation of this packet and further stop dequeue any following BH packets (line 6-7).

However, if the PW arrives late (e.g., $t_{n-1} > t_1 - T$ when PW is updated), there could be too many BH packets already queued on the WNIC, blocking the incoming LS packets. To avoid this problem, we design the synergy between the preemption controller and the virtual-preemption layer as follows (line 21): the controller updates the PW whenever a PW ends (i.e., $t_2 \leq T$) and notifies the virtual-preemption layer upon the update; the virtual-preemption layer first updates its PW to the latest and then resumes the dequeue operation of the BH flows on the driver. In this way, when the driver attempts to dequeue a BH packet, the virtual-preemption layer will always estimate whether the BH packet will block the LS flows on the WNIC according to the latest PW.

The CTT maintained in the virtual-preemption layer guarantees an accurate estimation of transmission completion time. Based on the accurate estimation, the virtual-preemption layer can minimize the time for which the driver stops transmitting BH packets (**R2**) while guaranteeing that BH packets queued on the WNIC will be finished transmitting before LS packets arrive (**R1**).

Algorithm 2: Virtual-Preemption Layer

Data: CTT: completion time table; $[t_1, t_2]$: current PW; N_{samp}: number of samples to keep in the CTT; pkt: BH packet

upon dequeue pkt of kerenl queue **do**
  $n \leftarrow$ current number of packets in hardware queue;
  $t_n \leftarrow max(CTT[n])$;
  $T \leftarrow now()$;
  if $(T < t_1$ && $t_1 - T \leq t_n)$ || $t_1 \leq T \leq t_2$ **then**
    cancel dequeue;
    stop kernel BH queue;
  else
    let the driver dequeue pkt;
  end
upon enqueue pkt of BH flow **do**
  $n \leftarrow$ current number of packets in hardware queue;
  $t_{enq} \leftarrow now()$;
  $pkt.t \leftarrow t_{enq}$;
  $pkt.n \leftarrow n$;
upon complete pkt of BH flow **do**
  $t_{comp} \leftarrow now()$; **if** $|CTT[pkt.n]| > N_{samp}$ **then**
    remove oldest sample from $CTT[pkt.n]$;
  end
  $CTT[pkt.n] \leftarrow CTT[pkt.n] \bigcup \{t_{comp} - pkt.t\}$;
upon receive PW $[t'_1, t'_2]$ **do**
  $t_1 \leftarrow t'_1$;
  $t_2 \leftarrow t'_2$;
  start kernel BH queue;

D. Theoretical Properties of COORP

Here we present two properties of COORP. For the simplicity of the analysis, we set the LIMIT of the global controller to 1. We define the latency of an LS message as the time between when it arrives at the WNIC and when it is received by another WNIC. First, we denote several latency distributions for the analysis:

D_{free}: the latency of the LS messages when there is no contention and local congestion.

D_{EDCA}: the latency of the LS messages when it is granted the highest priority in EDCA and contends with one BH flow.

D_{cong}: the latency of the LS messages when there is only local congestion.

We also use $D(x)$ to denote the cumulative distribution function (CDF) of D. Note that these latency distributions can be obtained with detailed theoretical analysis [44], [45] or with experiments. The following analysis is independent of how these distributions are obtained.

Property 1. In COORP, at different stages of CRL and on different robots, the latency CDFs of LS messages can follow three cases: $D_{free}(x)$ or $D_{EDCA}(x)$ or $pD_{free}(x) + (1-p)D_{cong}(x)$, where p is the possibility that the arrival time of LS messages falls in the corresponding PW.

Proof sketch. In COORP, BH flows are controlled by the global controller, and only one robot will be transmitting BH flows at a time with the LIMIT set to one. Note that we omit the influence of congestion and contention among different LS flows because they are small in size. There are three different states from a robot's perspective: 1) no one is transmitting BH flows; 2) the robot itself is transmitting BH flows; 3) another robot is transmitting BH flows. In case 1 and case 3, the CDF is $D_{free}(x)$ and $D_{EDCA}(x)$ respectively by definition. In case 2, due to the protection of COORP, portion p of the messages will get no congestion and follow D_{free}, portion $1-p$ of the messages will get local congestion and follow D_{cong}. Thus the latency CDF is the mixture [46] of $D_{free}(x)$ and $D_{cong}(x)$, with weight p and $1-p$ correspondingly, which is $pD_{free}(x) + (1-p)D_{cong}(x)$.

Property 2. The bandwidth utilization of BH flows in COORP is

$$1 - 2F \times norminv(\frac{1+C}{2}) \tag{1}$$

where $norminv(x)$ is the inverse of the cumulative distribution function for the fluctuations of LS message arrival time which follow $N(0, \sigma^2)$, C is the configured accuracy requirement as was described before, and F is the frequency of LS messages.

Proof sketch. First we determine the length of PW. $P(-\frac{PW}{2} < X < \frac{PW}{2}) = C, X \sim N(0, \sigma^2)$. Thus $P(X < \frac{PW}{2}) = \frac{1+C}{2}$, leading to $PW = 2 \times norminv(\frac{1+C}{2})$.

In COORP, the virtual-preemption layer pauses the BH flow for time PW with frequency F. Thus during each time slice with length L, the time paused totally is $L \times F \times PW$. So the bandwidth utilization is $\frac{L-L\times F\times PW}{L} = 1 - 2F \times norminv(\frac{1+C}{2})$.

V. IMPLEMENTATION

The virtual-preemption layer was implemented as a kernel device module with about 600 lines of code based on Linux 5.4.84 with PREEMP_RT patch [47], and interfaced with the preemption controller at user space through `ioctl()`. To get the completion time of each packet, the virtual-preemption layer was hooked into two operations of the driver. The operation to enqueue a packet to the WNIC queue was hooked to record the current number of enqueued packets in the WNIC queue and the timestamp. The operation to handle packet completion from the WNIC queue was hooked to compute the time that this packet and the previous enqueued packets take to be transmitted in the WNIC queue, and update the CTT as was described in Algorithm2. Further, the virtual-preemption layer was hooked into the driver operation of dequeueing packets from mac80211 queues to prevent it from enqueueing too many packets into the WNIC queue. Although the exact code of WNIC drivers differs, these three operations are general and necessary. For evaluation, the hooks were added into the MT76 driver in the kernel source tree with about 60 lines of modification, and the functions corresponding to the above three operations were `mt76_queue_ops.tx_queue_skb()`, `mt76_queue_ops.tx_complete_skb()` and `mt76_txq_dequeue()`.

We implemented the preemption controller as two decoupled modules, a proxy in the ROS2 rcl (ROS client library) [48] to compute the time model for LS flow, and a controller as a standalone ROS2 application. For the communication between the proxy and the controller, instead of using traditional inter-process communication methods like sockets and shared memory, we chose the ROS2 way, namely publishing the time model to a special topic that is subscribed by the controller. This made COORP modular and would enable developing more tools based on the information collected by COORP. It involved around 400 lines of code in the ROS2 rcl.

Ideally, the BH controller should also be implemented in ROS2. However, ROS2 does not currently provide direct TCP support, and the default UDP support not only limits the size of the message (64KB), but also suffers bandwidth problems in unreliable networks like WiFi [49].

Thus we chose to temporarily implement a message streaming interface with TCP socket, integrated with the BH controller. This also faces technical difficulties. Realizing the BH controller requires precisely pausing and resuming the actual transmission of the TCP socket, i.e., pausing the transmission at the end of each time slice and resuming the transmission at the start of each time slice. Unfortunately, the POSIX socket API does not provide such ability to pause/resume the actual transmission. To address this difficulty, we added a new TCP socket option `TCP_PAUSE` which can be controlled by calling `setsockopt()` [50]. To implement the semantics of `TCP_PAUSE`, we modified the TCP output engine [51] such that `tcp_write_xmit()` sends no packet when `TCP_PAUSE` is set to true. Note that this implementation does not interfere with the congestion control algorithm of TCP since the packets are not sent out from TCP's perspective.

VI. EVALUATION

Testbed. Our evaluation was done using 5 hosts running Ubuntu 20.04 with Linux 5.4.84 kernel patched with Preempt-RT [52]. Two of the hosts were desktops with Intel Core i7-8700 CPU @ 3.20GHz, and the other three were laptops with different models of CPUs, namely Intel Core i7-10510 CPU@1.80GHz, i5-7200U CPU@2.50GHz, and i7-7700HQ CPU @ 2.80GHz respectively. We simulate robots with these five hosts and refer to these hosts as robots for convenience in this section. One of them (the leader) was equipped with MediaTek MT7612E WNIC and configured to AP mode on channel 44 (5GHz). The other four were equipped with MediaTek MT7612U USB WNICs and were connected to the leader. For the measurement of latency, the hosts were connected to a TP-LINK TL-SG108 desktop Ethernet switch, and their system clocks were synchronized with PTP [53]. The synchronization accuracy was reported to be below 5μs by `ptp4l`, sufficient for measuring wireless transmission latency. Note that the accurate time synchronization is only for measuring the latency and reaction time, and is not required by COORP to work in real deployments.

Baselines. We compared COORP with two methods chosen from the two categories of existing work. We choose EDCA [17] from the prioritized contention category because it is a part of the IEEE 802.11 standard and is supported by commodity WNICs. We choose SchedWiFi [20] from the global planning category because it was specifically designed for periodic latency-sensitive traffic like perception messages in CRL. For EDCA, the LS messages were configured to go through `AC_VO`, the highest priority, and the training data and parameter updates were configured

to go through AC_BE, lower priority than the LS messages. These configurations are the same as COORP.

For SchedWiFi, since the original paper only evaluated with simulation, to compare its actual performance with COORP, we implemented all its features. ST-Window (STW) is the time window reserved for each LS flow; it was calculated through Equation 2 following the original paper, where L is the size of each message in an LS flow, $SIFS$ is $16\mu s$ according to the standards of IEEE 802.11 [54] and the maximum retransmission time R was set to 7 [55]. TX_{ack} was empirically set to $30\mu s$. The Time-Aware Shaper of SchedWiFi required modification to the PHY layer of the WNICs, which is impractical on commodity WNICs. To achieve the same effect, we adapted COORP's virtual-preemption layer as its Time-Aware Shaper.

$$STW = 2 \times 25\mu s + (\frac{L}{bandwidth} + 2 \times SIFS + TX_{ack}) \times (1 + R) \quad (2)$$

In the evaluation, we set the limit of COORP's global controller to 1 for the lowest violation rate of the reaction time boundary.

Workload. We adopted the cooperative multi-robot navigation task in [2] as the workload, since this task requires multiple robots to gather their real-time sensor data for cooperative control, and is an important building block of mission-critical multi-robot applications such as surveillance and rescue [2], [56], [57]. We set up the task in the Stage simulator with 5 moving obstacles and 5 robots. We used the same architecture of the deep reinforcement learning model described in [2]. Nowadays, the parameter size of many real-world reinforcement learning models for robotics has exceeded 100 million [58] and many methods can reduce the number of parameters by 10x-50x [59]. So we extended the number of parameters to 10 million to approximate the real-world reinforcement learning model for wireless distributed learning.

To bridge the simulated robots with real-world wireless network, the perception of the robots was first retrieved from the simulator and then sent to the corresponding host through the wired network. Each robot reported the perception it received to the leader via the wireless network as if the perception was collected from local sensors. The leader combined the perceptions and fed the combined perception to the model for inference. The leader then generated cooperative control messages and disseminated them to the corresponding robot via the wireless network.

Our evaluation focused on these questions:

- RQ1: How is COORP compared to baseline systems in terms of *fast reaction* and *fast model convergence*?
- RQ2: What is the effectiveness of different components of COORP?
- RQ3: How does COORP scale with the number of robots?
- RQ4: What are the limitations of COORP?

A. End-to-end Performance

For end-to-end comparison between COORP and the baseline systems, we recorded their reaction time (i.e., the time of MPC loops) during training and their model convergence speed (i.e., the score the model achieved after the learning process started for a certain time). Figure 2a shows the CDF of reaction time of different systems. The vertical line marks the typical reaction time boundary ($\frac{1}{33}s$) for mission-critical robotic applications. 'MPC loop only' corresponds to the reaction time when the learning process was not started and no BH flows were involved, which was the best reaction time any system can achieve with the same hardware and physical layer protocol.

Among all the three systems, EDCA had the highest violation rate (53.9%) of the reaction time boundary, since it did not resolve local congestion in the WNIC, and the simultaneous transmission of multiple (1 to 5) BH flows resulted in heavy global congestion and amplified local congestion. SchedWiFi best guaranteed fast reaction with a 3.5% violation rate and was very close to 'MPC loop only', because LS messages were assigned exclusive and long transmission time slots according to Equation 2, so as to cope with various uncertainty and avoid interference of any BH messages. COORP achieved the second lowest violation rate (8.8%) of the reaction time boundary and was also close to 'MPC loop only'. These results correspond to the recorded CDF of latency of LS messages achieved by different systems in Figure 7: COORP achieved the second lowest latency for LS messages, close to 'MPC loop only', and comparable to the lowest one, SchedWiFi.

The convergence of the control model during training is shown in Figure 2b (BSP) and Figure 8 (SSP, the limit of the number of batches between the fastest and the slowest worker was set to 1). All three systems started from the same pre-trained model. EDCA was the fastest one among these three systems in terms of model convergence. In contrast, SchedWiFi was 85.3% and xxx% slower despite its low violation rate of the reaction time boundary. COORP achieved both a low violation rate of the reaction time boundary and a minor slowdown of the training

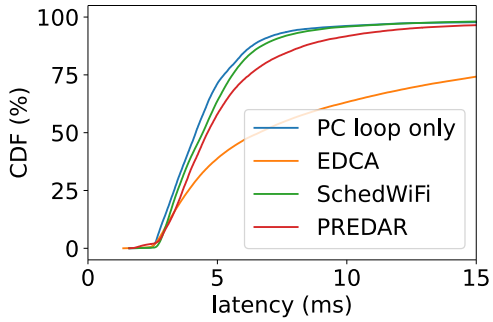


Fig. 7: Latency CDF of LS messages.

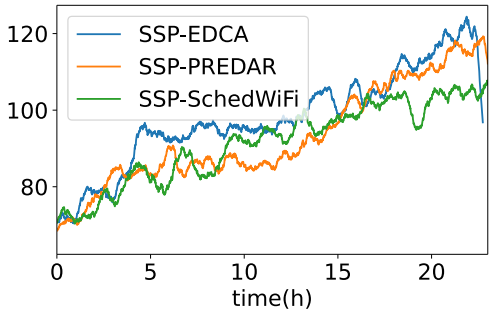


Fig. 8: Convergence over time, SSP.

TABLE I: Number of training iterations per hour, ASP.

	EDCA	SchedWiFi	COORP
BSP	4.15	2.77	3.60
SSP	4.41	2.82	3.88
ASP	4.61	2.95	3.96

process. We found ASP not converging for this task using all the three systems, thus we only show its number of training iterations finished per hour as Table I. The non-convergence of ASP would be due to outdated updates from slow workers [30].

TABLE II: Comparison of bandwidth utilization. Normalized to EDCA's.

	EDCA	SchedWiFi	COORP
utilization	1.00 (194.7Mbps)	0.59 (115.1Mbps)	0.86 (168.2Mbps)

Table II shows the bandwidth utilization during the learning process of different systems. EDCA achieved the highest throughput (i.e. bytes transmitted per second), and we treated its throughput as the bandwidth of the underlying hardware and physical layer protocol. The bandwidth utilization in SchedWiFi dropped by 41% since all the robots had to pause BH flows for all the time slots for LS messages, and the sizes of the time slots were often overestimated, limiting the overall time for BH flows to transmit. In COORP, although only one robot was allowed to transmit BH flow at any time, the only transmitting BH flow achieved almost the same bandwidth utilization as multiple BH flows concurrently transmitting, and each robot only needed to pause for its own LS messages. Besides, the coordination between global controller and

BH controller incurred little overhead. Thus there was only a minor drop (14%) of the throughput in COORP. The higher bandwidth utilization in COORP led to faster model convergence compared to SchedWiFi, shown in Figure 2b and Figure 8.

B. *Effectiveness of Components*

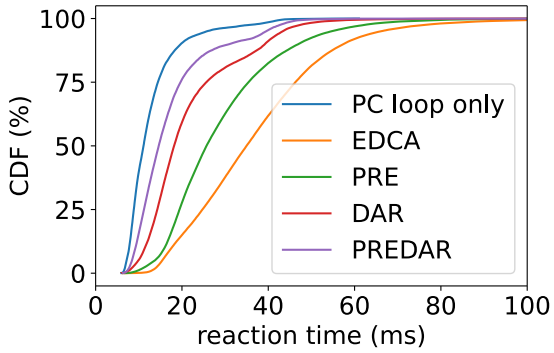


Fig. 9: Contribution of COORP's components to the reaction time. In 'PRE', only preemption was enabled. In 'DAR', only DP-aware contention reduction was enabled.

COORP consists of two key components: preemption and DP-aware contention reduction. To understand the effectiveness of the two components, we enabled each of the two components individually and measured its contribution to the reduction of reaction time, as shown in Figure 9. When DP-aware contention reduction was enabled alone, not only had the global congestion been eliminated, but the local congestion was no longer amplified. When preemption was enabled alone, local congestion was eliminated, but global congestion remains. Thus DP-aware contention reduction was more effective when enabled alone, but the combination of them eliminated both global and local congestions, and led to lower reaction time.

To further validate the effectiveness of the virtual-preemption layer, we inspected the number of BH packets queued in the WNIC when LS packets were delivered to the WNIC, shown in Figure 10. In EDCA, a large number of packets from BH flows accumulated in the queue of WNICs and would block the transmission of LS packets. In COORP, the virtual-preemption layer avoided delivering too many BH packets to the WNIC before LS packets. Thus, the WNIC was able to finish transmitting BH packets before LS packets were delivered and transmit LS packets as soon as possible. However, since the virtual-preemption layer infers the transmission completion time based on historical records, it was unable to foresee jitters of the transmission

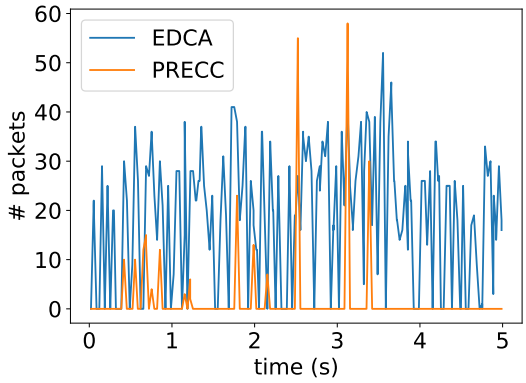


Fig. 10: Effectiveness of the virtual-preemption layer. Y-axis is the number of packets in the WNIC queue when LS packets enqueue, which is zero most of the time in COORP.

and could not ensure the WNIC queue was always clean. As was discussed in Section II, a solution like the Time-Aware Shaper of SchedWiFi integrated within the WNICs would be the ultimate solution to local congestion, but it is still unavailable due to the required efforts from both wireless network standard and WNIC vendors. The virtual-preemption layer of COORP would serve as a universal and effective solution.

C. Scalability to the Number of Robots

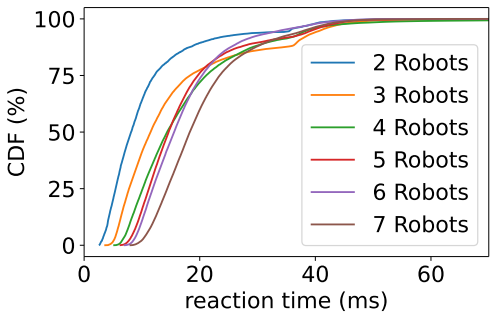


Fig. 11: Scalability of reaction time.

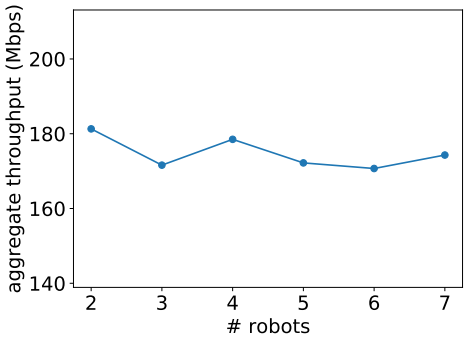


Fig. 12: Scalability of bandwidth utilization.

Figure 11 shows the scalability of COORP in terms of reaction time. The increase of reaction time comes from two aspects. First, the contention among LS messages from different robots

increases latency. Second, with more robots, it takes more time to gather all the perceptions from all the robots.

In COORP, the overhead for controlling contention comes from the requests and responses between BH controller and global controller, which is small compared to the transmission of parameter updates. The preemption requires each robot only to pause for its own LS messages, which is independent of the number of robots. Thus, the bandwidth utilization scaled well as shown in Figure 12, and the differences mainly come from random jitters.

D. Performance under Low Available Bandwidth

In CRL, robots move around and may result in a drop in available bandwidth. To understand the performance of COORP under such scenarios, we set up two individual cases.

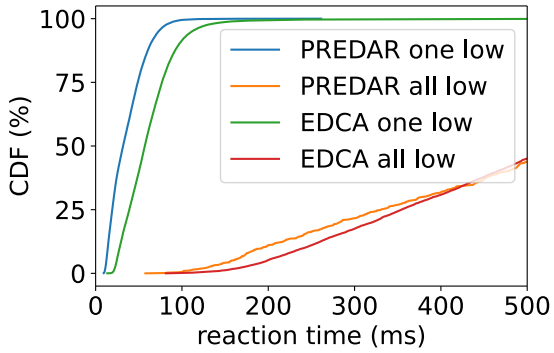


Fig. 13: Reaction time under low available bandwidth.

TABLE III: Bandwidth utilization under low available bandwidth. In 'one(all) low', one(all) of the robots used the lowest transmission rate.

	EDCA	COORP
one low	99.8Mbps	148.1Mbps
all low	19.6Mbps	16.3Mbps

In the first case, denoted as 'one low', one of the robots was configured to use the lowest transmission rate `vht-mcs-5 1:0` using `iw` tool. Compared to COORP, EDCA achieved a lower throughput. This would be related to the performance anomaly of 802.11 previously discussed by [37], [60] that the stations in a wireless network would achieve similar throughput even when

their network conditions differ. By limiting the number of robots to transmit simultaneously (1 robot in the evaluation), COORP enabled each robot to achieve the highest throughput allowed by its network condition, resulting in higher aggregate throughput. This indicates that COORP provides faster training than EDCA when the available bandwidth of a small number of robots is low.

In the second case, denoted as 'all low', all the robots were configured to use the lowest transmission rate. The throughput was restricted by the physical connection, thus both EDCA and COORP suffered a low throughput which would inevitably slow down the training.

E. Limitations

COORP has the following limitations. First, COORP focused on the case of a single robot group and did not handle the case when multiple robot groups coexist and use the same wireless channel. We leave this as our future work. Second, instead of real robots, COORP was evaluated on a semi-simulated testbed where the interaction between the robots and the environment was simulated while the communication among the robots goes through a real-world wireless network. Since this paper focused on the network supporting systems, such settings would be sufficient for the evaluation purpose.

VII. CONCLUSION

In this paper, we identify two technical requirements of the network supporting system for mission-ciritcal coordinated robotic learning, namely fast reaction and fast model convergence, and present COORP, the first network supporting system that achieves both of the requirements. COORP enables robot groups to adapt to new environments with minimal time consumption while avoiding lag and collisions. COORP's code is released on github.com.

REFERENCES

[1] J. P. Queralta, J. Taipalmaa, B. Can Pullinen, V. K. Sarker, T. Nguyen Gia, H. Tenhunen, M. Gabbouj, J. Raitoharju, and T. Westerlund, "Collaborative multi-robot search and rescue: Planning, coordination, perception, and active vision," vol. 8, pp. 191 617–191 643, publisher: IEEE.

[2] R. Han, S. Chen, and Q. Hao, "Cooperative Multi-Robot Navigation in Dynamic Environment with Deep Reinforcement Learning," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, May 2020, pp. 448–454, iSSN: 2577-087X.

[3] E. Vidal, J. D. Hernández, N. Palomeras, and M. Carreras, "Online robotic exploration for autonomous underwater vehicles in unstructured environments," in *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*, pp. 1–4.

[4] Y. Zhang, D. Shi, Y. Wu, Y. Zhang, L. Wang, and F. She, "Networked Multi-robot Collaboration in Cooperative–Competitive Scenarios Under Communication Interference," in *Collaborative Computing: Networking, Applications and Worksharing*, H. Gao, X. Wang, M. Iqbal, Y. Yin, J. Yin, and N. Gu, Eds. Cham: Springer International Publishing, 2021, vol. 349, pp. 601–619. [Online]. Available: http://link.springer.com/10.1007/978-3-030-67537-0_36

[5] R. Nishihara, P. Moritz, S. Wang, A. Tumanov, W. Paul, J. Schleier-Smith, R. Liaw, M. Niknami, M. I. Jordan, and I. Stoica, "Real-Time Machine Learning: The Missing Pieces," *arXiv:1703.03924 [cs]*, May 2017, arXiv: 1703.03924. [Online]. Available: http://arxiv.org/abs/1703.03924

[6] P. Caloud, Wonyun Choi, J. Latombe, C. L. Pape, and M. Yim, "Indoor automation with many mobile robots," in *EEE International Workshop on Intelligent Robots and Systems, Towards a New Frontier of Applications*, pp. 67–72 vol.1.

[7] J. L. Baxter, E. K. Burke, J. M. Garibaldi, and M. Norman, "Multi-robot search and rescue: A potential field based approach," in *Autonomous Robots and Agents*, ser. Studies in Computational Intelligence, S. C. Mukhopadhyay and G. S. Gupta, Eds. Springer, pp. 9–16. [Online]. Available: https://doi.org/10.1007/978-3-540-73424-6_2

[8] X. Kong, B. Xin, F. Liu, and Y. Wang, "Revisiting the master-slave architecture in multi-agent deep reinforcement learning." [Online]. Available: http://arxiv.org/abs/1712.07305

[9] K. Zhang, Z. Yang, and T. Basar, "Networked multi-agent reinforcement learning in continuous spaces," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 2771–2776.

[10] H. Nascimento, M. Mujica, and M. Benoussaad, "Collision avoidance interaction between human and a hidden robot based on kinect and robot data fusion," *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 88–94, 2021, publisher: IEEE. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03139189

[11] X. Sun, X. Weng, and K. Kitani, "When We First Met: Visual-Inertial Person Localization for Co-Robot Rendezvous," *arXiv:2006.09959 [cs]*, Nov. 2020, arXiv: 2006.09959. [Online]. Available: http://arxiv.org/abs/2006.09959

[12] L. S. Scimmi, M. Melchiorre, S. Mauro, and S. Pastorelli, "Experimental Real-Time Setup for Vision Driven Hand-Over with a Collaborative Robot," in *2019 International Conference on Control, Automation and Diagnosis (ICCAD)*, Jul. 2019, pp. 1–5.

[13] D. Kim, I. Yeom, and T.-J. Lee, "Mitigating tail latency in IEEE 802.11–based networks," *International Journal of Communication Systems*, vol. 31, no. 1, p. e3404, 2018, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/dac.3404. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/dac.3404

[14] Changhua Pei, Youjian Zhao, Yunxin Liu, Kun Tan, Jiansong Zhang, Yuan Meng, and Dan Pei, "Latency-based WiFi congestion control in the air for dense WiFi networks," in *2017 IEEE/ACM 25th International Symposium on Quality of Service (IWQoS)*, Jun. 2017, pp. 1–10.

[15] M. Heusse, F. Rousseau, R. Guillier, and A. Duda, "Idle Sense: An Optimal Access Method for High Throughput and Fairness in Rate Diverse Wireless LANs," p. 12.

[16] A. Saeed, M. Ammar, E. Zegura, and K. Harras, "If you can't Beat Them, Augment Them: Improving Local WiFi with Only Above-Driver Changes," in *2018 IEEE 26th International Conference on Network Protocols (ICNP)*, Sep. 2018, pp. 378–388, iSSN: 1092-1648.

[17] "IEEE 802.11e-2005," Feb. 2021, page Version ID: 1006381497. [Online]. Available: https://en.wikipedia.org/w/index.php?title=IEEE_802.11e-2005&oldid=1006381497

[18] Z. Yang, J. Zhang, K. Tan, Q. Zhang, and Y. Zhang, "Enabling TDMA for today's wireless LANs," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, Apr. 2015, pp. 1436–1444, iSSN: 0743-166X.

[19] Y. Wei, Q. Leng, S. Han, A. K. Mok, W. Zhang, and M. Tomizuka, "RT-WiFi: Real-Time High-Speed Communication Protocol for Wireless Cyber-Physical Control Applications," in *2013 IEEE 34th Real-Time Systems Symposium*, Dec. 2013, pp. 140–149, iSSN: 1052-8725.

[20] G. Patti, G. Alderisi, and L. L. Bello, "SchedWiFi: An innovative approach to support scheduled traffic in ad-hoc industrial IEEE 802.11 networks," in *2015 IEEE 20th Conference on Emerging Technologies Factory Automation (ETFA)*, Sep. 2015, pp. 1–9, iSSN: 1946-0759.

[21] "IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks–Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016)*, pp. 1–4379, Feb. 2021, conference Name: IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016).

[22] "Frame aggregation," Sep. 2020, page Version ID: 978836459. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Frame_aggregation&oldid=978836459

[23] "ROS Documentation." [Online]. Available: https://docs.ros.org/

[24] J. Scherer, S. Yahyanejad, S. Hayat, E. Yanmaz, T. Andre, A. Khan, V. Vukadinovic, C. Bettstetter, H. Hellwagner, and B. Rinner, "An Autonomous Multi-UAV System for Search and Rescue," in *Proceedings of the First Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use*. Florence Italy: ACM, May 2015, pp. 33–38. [Online]. Available: https://dl.acm.org/doi/10.1145/2750675.2750683

[25] R. N. Darmanin and M. K. Bugeja, "A review on multi-robot systems categorised by application domain," in *2017 25th Mediterranean Conference on Control and Automation (MED)*, Jul. 2017, pp. 701–706, iSSN: 2473-3504.

[26] M. J. Matarić, "Reinforcement Learning in the Multi-Robot Domain," in *Robot Colonies*, R. C. Arkin and G. A. Bekey, Eds. Boston, MA: Springer US, 1997, pp. 73–83. [Online]. Available: https://doi.org/10.1007/978-1-4757-6451-2_4

[27] J. Ibarz, J. Tan, C. Finn, M. Kalakrishnan, P. Pastor, and S. Levine, "How to train your robot with deep reinforcement learning: lessons we have learned," *The International Journal of Robotics Research*, vol. 40, no. 4-5, pp. 698–721, Apr. 2021, publisher: SAGE Publications Ltd STM. [Online]. Available: https://doi.org/10.1177/0278364920987859

[28] S. Dasari, F. Ebert, S. Tian, S. Nair, B. Bucher, K. Schmeckpeper, S. Singh, S. Levine, and C. Finn, "RoboNet: Large-Scale Multi-Robot Learning," *arXiv:1910.11215 [cs]*, Jan. 2020, arXiv: 1910.11215. [Online]. Available: http://arxiv.org/abs/1910.11215

[29] A. Nagabandi, I. Clavera, S. Liu, R. S. Fearing, P. Abbeel, S. Levine, and C. Finn, "Learning to Adapt in Dynamic, Real-World Environments Through Meta-Reinforcement Learning," *arXiv:1803.11347 [cs, stat]*, Feb. 2019, arXiv: 1803.11347. [Online]. Available: http://arxiv.org/abs/1803.11347

[30] X. Zhao, A. An, J. Liu, and B. X. Chen, "Dynamic Stale Synchronous Parallel Distributed Training for Deep Learning," *arXiv:1908.11848 [cs, stat]*, Aug. 2019, arXiv: 1908.11848. [Online]. Available: http://arxiv.org/abs/1908.11848

[31] F. Santos, L. Almeida, and L. S. Lopes, "Self-configuration of an adaptive TDMA wireless communication protocol for teams of mobile robots," in *2008 IEEE International Conference on Emerging Technologies and Factory Automation*, Sep. 2008, pp. 1197–1204, iSSN: 1946-0759.

[32] S. Cao and V. C. S. Lee, "A Novel Adaptive TDMA-Based MAC Protocol for VANETs," *IEEE Communications Letters*, vol. 22, no. 3, pp. 614–617, Mar. 2018, conference Name: IEEE Communications Letters.

[33] X. Guo, S. Wang, H. Zhou, J. Xu, Y. Ling, and J. Cui, "Performance Evaluation of the Networks with Wi-Fi based TDMA Coexisting with CSMA/CA," *Wireless Personal Communications*, vol. 114, no. 2, pp. 1763–1783, Sep. 2020. [Online]. Available: https://doi.org/10.1007/s11277-020-07447-3

[34] D. Taht, J. Gettys, E. Dumazet, T. Hoeiland-Joergensen, E. Dumazet, P. McKenney, J. Gettys, T. Hoeiland-Joergensen, and P. McKenney, "The Flow Queue CoDel Packet Scheduler and Active Queue Management Algorithm." [Online]. Available: https://tools.ietf.org/html/rfc8290

[35] S. Seytnazarov and Y. Kim, "QoS-aware adaptive MPDU aggregation of VoIP traffic on IEEE 802.11n WLANs," in *10th*

International Conference on Network and Service Management (CNSM) and Workshop, Nov. 2014, pp. 356–359, iSSN: 2165-963X.

[36] T. Høiland-Jørgensen, P. Hurtig, and A. Brunstrom, "The Good, the Bad and the WiFi: Modern AQMs in a residential setting," *Computer Networks*, vol. 89, pp. 90–106, Oct. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1389128615002479

[37] T. Høiland-Jørgensen, M. Kazior, D. Täht, A. Brunstrom, and P. Hurtig, "Ending the Anomaly: Achieving Low Latency and Airtime Fairness in WiFi," p. 15, 2017.

[38] C. A. Grazia, "Towards 1 ms WLAN Latency," in *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct. 2018, pp. 278–284, iSSN: 2160-4886.

[39] A. Showail and B. Shihada, "Battling Latency in Modern Wireless Networks," *IEEE Access*, vol. 6, pp. 26 131–26 143, 2018, conference Name: IEEE Access.

[40] M. Backhaus, M. Theil, M. Rossberg, and G. Schaefer, "Towards a Flexible User-Space Architecture for High-Performance IEEE 802.11 Processing," in *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct. 2018, pp. 1–9, iSSN: 2160-4886.

[41] D. Tardioli, R. Parasuraman, and P. Ögren, "Pound: A multi-master ROS node for reducing delay and jitter in wireless multi-robot networks," *Robotics and Autonomous Systems*, vol. 111, pp. 73–87, Jan. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889017309144

[42] "IEEE 802.11 RTS/CTS," Nov. 2020, page Version ID: 991386765. [Online]. Available: https://en.wikipedia.org/w/index.php?title=IEEE_802.11_RTS/CTS&oldid=991386765

[43] "Hidden node problem," Oct. 2020, page Version ID: 984728511. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Hidden_node_problem&oldid=984728511

[44] P. Schulz, L. Ong, P. Littlewood, B. Abdullah, M. Simsek, and G. Fettweis, "End-to-End Latency Analysis in Wireless Networks with Queuing Models for General Prioritized Traffic," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2019, pp. 1–6, iSSN: 2474-9133.

[45] K. Medepalli and F. A. Tobagi, "On Optimization of CSMA/CA based Wireless LANs: Part I - Impact of Exponential Backoff," in *2006 IEEE International Conference on Communications*, vol. 5, Jun. 2006, pp. 2089–2094, iSSN: 1938-1883.

[46] "Mixture distribution," Mar. 2021, page Version ID: 1013174298. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Mixture_distribution&oldid=1013174298

[47] RTwiki. [Online]. Available: https://rt.wiki.kernel.org/index.php/Main_Page

[48] "ros2/rcl," Apr. 2021, original-date: 2015-02-21T00:06:40Z. [Online]. Available: https://github.com/ros2/rcl

[49] "DDS tuning information." [Online]. Available: https://index.ros.org/doc/ros2/Troubleshooting/DDS-tuning/

[50] "getsockopt(2) - Linux manual page." [Online]. Available: https://man7.org/linux/man-pages/man2/getsockopt.2.html

[51] "How the Linux TCP output engine works." [Online]. Available: http://vger.kernel.org/~davem/tcp_output.html

[52] "realtime:start [Wiki]." [Online]. Available: https://wiki.linuxfoundation.org/realtime/start

[53] "Precision Time Protocol," Jan. 2021, page Version ID: 1001872854. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Precision_Time_Protocol&oldid=1001872854

[54] "Short interframe space," page Version ID: 988080746. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Short_Interframe_Space&oldid=988080746

[55] "802.11/MAC/Lower/Retransmissions – WARP Project." [Online]. Available: https://warpproject.org/trac/wiki/802.11/MAC/Lower/Retransmissions

[56] P. Long, T. Fan, X. Liao, W. Liu, H. Zhang, and J. Pan, "Towards Optimally Decentralized Multi-Robot Collision

Avoidance via Deep Reinforcement Learning," *arXiv:1709.10082 [cs]*, May 2018, arXiv: 1709.10082. [Online]. Available: http://arxiv.org/abs/1709.10082

[57] A. Benevento, M. Santos, G. Notarstefano, K. Paynabar, M. Bloch, and M. Egerstedt, "Multi-Robot Coordination for Estimation and Coverage of Unknown Spatial Fields," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, May 2020, pp. 7740–7746, iSSN: 2577-087X.

[58] "RoboNet: A Dataset for Large-Scale Multi-Robot Learning – The Berkeley Artificial Intelligence Research Blog." [Online]. Available: https://bair.berkeley.edu/blog/2019/11/26/robo-net/

[59] N. Liu, X. Ma, Z. Xu, Y. Wang, J. Tang, and J. Ye, "AutoCompress: An automatic DNN structured pruning framework for ultra-high compression rates," vol. 34, no. 4, pp. 4876–4883, number: 04. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/5924

[60] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*, vol. 2, Mar. 2003, pp. 836–843 vol.2, iSSN: 0743-166X.