



# The viability of domain constrained coalition formation for robotic collectives

Grace Diehl<sup>1</sup> · Julie A. Adams<sup>1</sup>

Received: 4 February 2023 / Accepted: 15 June 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

Applications, such as military and disaster response, can benefit from robotic collectives' ability to perform multiple cooperative tasks (e.g., surveillance, damage assessments) efficiently across a large spatial area. *Coalition formation* algorithms can potentially facilitate collective robots' assignment to appropriate task teams; however, most coalition formation algorithms were designed for smaller multiple robot systems (i.e., 2–50 robots). Collectives' scale and domain-relevant constraints (i.e., distribution, near real-time, minimal communication) make coalition formation more challenging. This manuscript identifies the challenges inherent to designing coalition formation algorithms for very large collectives (e.g., 1000 robots). A survey of multiple robot coalition formation algorithms finds that most are unable to transfer directly to collectives, due to the identified system differences; however, auctions and hedonic games may be the most transferable. A simulation-based evaluation of five total algorithms from two combinatorial auction families and one hedonic game family, applied to homogeneous and heterogeneous collectives, demonstrates that there are collective compositions for which no evaluated algorithm is viable; however, the experimental results and literature survey suggest paths forward.

**Keywords** Robotic collectives · Multiple robot systems · Coalition formation · Task allocation

## 1 Introduction

Robotic collectives, defined here as  $> 50$  robots with limited sensing and computational capabilities as compared to a typical robot in a smaller multiple robot system, are increasingly relevant for military (e.g., building or area surveillance) (Defense Advanced Research Projects Agency, 2019) and disaster response (e.g., damage assessments, search and rescue) (Hildmann & Kovacs, 2019) applications. A fundamental problem is *coalition*

---

✉ Grace Diehl  
diehlg@oregonstate.edu

Julie A. Adams  
julie.a.adams@oregonstate.edu

<sup>1</sup> Collaborative Robots and Intelligent Systems Institute, Oregon State University, Corvallis, OR 97331, USA

*formation for task allocation*, or partitioning robots into teams for task performance (Gerkey & Mataric, 2004; Korsah et al., 2013). Effective coalition formation can enable collectives to perform multiple cooperative tasks distributed over large environments efficiently; however, collectives' scale and the application domains' constraints make this problem challenging.

Military and disaster response applications require high-quality coalition formation solutions for hundreds to thousands of robots in near real-time (e.g., < 5 min). Low-bandwidth, deployed networks (e.g., mobile ad hoc networks) will often be utilized, due to damaged permanent infrastructure or remote deployment locations (Klinsompus & Nupairoj, 2015). Minimal, distributed communication is required to reserve bandwidth for mission critical communications (Jahir et al., 2019; Legendre et al., 2011; Shah et al., 2019). Additionally, frequent communication with a central entity is infeasible, as collectives can be highly distributed (Berman et al., 2009a; Hamann, 2018).

Many prior collective coalition formation algorithms assign approximate fractions of collectives to tasks (e.g., Berman et al., 2009b; Mather & Hsieh, 2011), or determine when robots participate in continuous tasks (e.g., Castello et al., 2013; Pang et al., 2019). These algorithms have been applied to large collectives, but do not address common robot coalition formation requirements, including optimizing solutions based on task utilities (i.e., numbers representing how important tasks are to a mission) and allowing team size specification. Many tasks require a minimum team size (e.g., cordon, area coverage), or an exact team size (e.g., moving a piece of debris, guarding a building entrance). This manuscript's focus is exact team sizes, as assigning more than the minimum robots unnecessarily reduces the resources available for new tasks, which can emerge frequently in highly dynamic domains.

Other existing coalition formation incorporates exact and approximation (e.g., Aziz et al., 2021; Dutta & Asaithambi, 2019; Service & Adams, 2011a; Zhang & Parker, 2013), auction-based (e.g., Chen & Sun, 2011; Guerrero et al., 2017; Oh et al., 2017), biologically-inspired (e.g., Agarwal et al., 2014; Mouradian et al., 2017; Yeh & Sugawara, 2016; Haque et al., 2013), and, recently, hedonic game-based (e.g., Czarnecki & Dutta, 2019; Jang et al., 2019) algorithms. However, software agent algorithms (e.g., Liemhetcharat & Veloso, 2014; Michalak et al., 2010; Rahwan & Jennings, 2008; Rahwan et al., 2009; Shehory & Kraus, 1995; Sless et al., 2014) are not directly transferable to collectives due to differences between software agents and embodied robots (Vig & Adams, 2007). Additionally, robot coalition formation algorithms were evaluated predominantly for smaller multiple robot systems, not collectives, and few evaluations consider practical communication requirements, so it is unknown if the algorithms will scale. Thus, existing algorithms either are not, or have not been demonstrated to be, viable for collectives.

This manuscript assesses robot coalition formation algorithms' viability. A simulation-based evaluation with up to one thousand robots demonstrates that multiple robot coalition formation algorithms do not fully satisfy the target domains' solution quality, runtime, and communication requirements when applied to collectives. In fact, there are collective compositions for which no existing algorithm is suitable; however, potential avenues for addressing these collective compositions are suggested.

## 1.1 Coalition formation for task allocation

Coalition formation for task allocation partitions agents into teams for task performance (Service & Adams, 2011b). The input is a set of  $n$  agents,  $A = \{a_1, \dots, a_n\}$ , a set of  $m$

tasks,  $T = \{t_1, \dots, t_m\}$ , and a *characteristic function*,  $f : (t_j, C_j) \rightarrow \mathbb{R}$ , where  $C_j \subseteq A$  is the set of agents assigned to task  $t_j$  (Service & Adams, 2011a). Characteristic functions vary by application; however,  $f(t_j, C_j)$  generally represents the inherent value of coalition  $C_j$  completing task  $t_j$  (Service & Adams, 2011a; Zhang & Parker, 2013).

A coalition formation for task allocation problem assumes that a coalition's value can be calculated considering only agents in the coalition (i.e., not considering agents assigned to other tasks) (Rahwan et al., 2009). The optimal solution is a *coalition structure*,  $CS = \{(t_1, C_1), \dots, (t_m, C_m)\}$ , that satisfies:

$$CS^* = \operatorname{argmax}_{CS} \sum_{j=1}^m f(t_j, C_j).$$

This manuscript considers a characteristic function where  $f(t_j, C_j)$  is equal to task  $t_j$ 's utility (i.e., representing the inherent value to the overall mission of completing  $t_j$ ) if  $C_j$  provides all of  $t_j$ 's required capabilities, and 0 otherwise. Maximizing the utility enables the collective to maximize its achievement of mission objectives; thus, the goal is for a collective coalition formation to achieve as close to 100% of the total mission utility as possible. However, deriving an optimal solution, or a provably reasonable approximation, is NP-hard (Service & Adams, 2011a; Sandholm et al., 1999). Specifically, no polynomial time algorithm can produce an  $O(|C|^{1-\epsilon})$  or  $O(m^{1-\epsilon})$  approximation, where  $C$  is the set of non-zero valued coalitions and  $\epsilon > 0$ , unless  $P = NP$  (Service & Adams, 2011a; Sandholm et al., 1999). Additionally, the establishment of any bound on solution quality requires considering at least  $O(2^{n-1})$  coalition structures (Sandholm et al., 1999). This computational complexity hinders producing high-utility coalition structures for large numbers of agents.

## 2 Coalition formation for multiple versus collective robot systems

Coalition formation has received much attention in the software multi-agent and multiple robot communities, where a *software multi-agent system* comprises two or more software agents, and a *multiple robot system* is two to fifty robots. Software multi-agent coalition formation cannot be transferred directly to robotic systems, as robots have different constraints and practical considerations (Vig & Adams, 2007). As embodied agents, robots have kinematic and dynamic constraints and relatively static, nontransferable capabilities (i.e., sensors and actuators) (Vig & Adams, 2007). Robots are also likely to have power constraints (Diehl & Adams, 2021b) and less available communication (Diehl & Adams, 2021a), which coalition formation must accommodate.

Multiple robot coalition formation considers embodied agents; however, transferring algorithms to robotic collectives can be complicated by differences in scale, capabilities, and communication. Key differences between collective and multiple robot systems must be considered when designing collective coalition formation algorithms, as summarized in Table 1.

### 2.1 Agent architecture

Robotic collectives typically incorporate simple control models (Brambilla et al., 2013) (e.g., repulsion, attraction, orientation (Hartman & Beneš, 2006), while multiple robot system architectures are often more sophisticated (e.g., belief-desire-intention). A reason for

**Table 1** Comparison of multiple robot and robotic collective systems

	Multiple robot system	Robotic collective
Definition	2–50 robots	> 50 robots
Agent architecture	Sophisticated control	Simple control
System capabilities	Emergent behavior possible	Emergent behavior by design
System composition	Homogeneous or heterogeneous	Homogeneous or reduced heterogeneity
Communication	Fully connected network	Location-dependent
	Rich messages	Small, simple messages

this difference is design choice. Robotic collectives were inspired by biological collectives, which achieve complex tasks through cooperation, rather than individual sophistication (Brambilla et al., 2013). Another reason is scale. Robotic collectives have more robots, so the allowable monetary and time costs per robot are lower; thus, even as robot technology and collective design improve, collective robots are likely to be less capable than their multiple robot system counterparts and have less computation available for coalition formation.

## 2.2 System capabilities

*System capabilities* are the tasks that a multiple robot system or collective can perform as a product of the individual robots' capabilities and interactions with other robots and the environment. Capabilities are relatively static when they correspond to hardware (e.g., sensors and actuators) (Vig & Adams, 2007), but *emergent behaviors* can vary dynamically (Beni, 2005).

Multiple robot systems can exhibit emergent behaviors, especially as more advanced artificial intelligence is incorporated (e.g., Costa et al., 2019); however, emergent behavior is intrinsic to collectives. Biologically-inspired collective capabilities (e.g., target selection Reina et al., 2015) typically rely on emergent behaviors, and the use of reactive architectures (see Sect. 2.1) renders interactions with the environment and other robots especially impactful (Hartman & Beneš, 2006). Collective coalition formation must adapt task assignments quickly as emergent capabilities change.

## 2.3 System composition

Collective and multiple robot systems may be *homogeneous* (i.e., robots have identical capabilities) or *heterogeneous* (i.e., some robots have different capabilities). Collective research has primarily considered homogeneous systems (e.g., Reina et al., 2015; Prorok et al., 2017; Van Der Blom & Bäck, 2018), while multiple robot coalition formation has considered heterogeneity extensively (e.g., Guerrero et al., 2017; Sen & Adams, 2013b; Zhang & Parker, 2012). Heterogeneous collectives enable more diverse applications (Defense Advanced Research Projects Agency, 2019; Clark et al., 2021; Prabhakar et al., 2020); thus, collective coalition formation must also account for heterogeneity.

Multiple robot systems can be highly heterogeneous, as each robot may be manufactured with unique capabilities (Hamann, 2018), while collectives' larger scale means that robots are likely to be manufactured in batches, where each robot in a given batch has the

same set of capabilities. Collective coalition formation algorithms may be able to leverage this reduced heterogeneity to decrease the required computation (Service & Adams, 2011a).

## 2.4 Communication

Coalition formation often assumes *fully connected* communication networks in which each robot can communicate with every other robot directly (e.g., Jang et al., 2018; Tang, 2006). This assumption can be feasible for multiple robot systems, if the deployment domains have good network coverage (Diehl & Adams, 2021b); however, robotic collectives, like their biological counterparts, are likely to rely on distributed, local communication to facilitate scalability (Berman et al., 2009a; Hamann, 2018). Individuals in *spatial swarms* (e.g., bird flocks (Ballerini et al., 2008), fish schools (Couzin et al., 2005)) often communicate with only a subset of their neighbors (Ballerini et al., 2008; Haque et al., 2016; Couzin et al., 2002; Strandburg-Peshkin et al., 2013), while *colonies* (e.g., bees (Seeley, 2010), ants (Gordon, 1999)) generally have fully connected communication topologies within a central hub and no, limited, or spatial swarm-like, communication outside the hub (Reina et al., 2015). Spatial swarms and colonies can theoretically approximate fully connected topologies by using local communication to propagate messages, but at the cost of delays and solution quality degradation; thus, collective robot coalition formation needs to rely primarily on local communication.

Many collective and multiple robot system applications will use temporary, deployable communication networks (e.g., ad hoc networks), due to damaged permanent infrastructure or remote deployment locations (Jahir et al., 2019; Legendre et al., 2011; Shah et al., 2019). Temporary network nodes (e.g., ground and aerial robots, satellites, balloons, blimps, tablets, or smart phones (Jahir et al., 2019; Pandey & De, 2017)) cannot replace permanent infrastructure fully, given their comparatively limited power and bandwidth (Muralidhar & Madhavi, 2018). Thus, multiple robot systems may send only medium-sized, rich messages, while collective robots may send only small, simple messages, due to the collectives' scale (e.g., thousands of robots, large area coverage). Collective coalition formation will need to accommodate limited message sizes.

## 3 Background and related work

Collective robots and coalition formation have, until recently (Dutta et al., 2021; Czarnecki & Dutta, 2021), been treated as distinct research areas, leaving collective coalition formation relatively unstudied. Collective and multiple robot coalition formation algorithms are discussed, as well as additional background on coalition formation.

### 3.1 The services model

Coalition formation requires determining if a coalition is capable of completing a task, given the coalition members' capabilities (Vig & Adams, 2007). There is a longstanding distinction between robot and software agent capability models, as hardware capabilities are not instantaneously transferable (Vig & Adams, 2007). Robots' software capabilities can also be more difficult to transfer, as software may be hardware-dependent (e.g.,

sensor-dependent controllers), and only limited communication may be available if a transfer is viable.

The robot *services model* considers high-level robot behaviors (e.g., surveillance) called *services* (Service & Adams, 2011a; Vig & Adams, 2007). Each robot has a service vector specifying the services that it can perform, and each task has a vector specifying the number of robots required to perform each service. A coalition can perform a task if there are sufficient robots to provide all required services, where each robot performs a single service at a time (Service & Adams, 2011a). This model is efficient when different sensors or actuators produce the same behavior, which supports scalability (Service & Adams, 2011a). Additionally, the services abstraction facilitates task design by users who are less familiar with robot hardware (e.g., tacticians (Defense Advanced Research Projects Agency, 2019)).

Three other common robot capabilities models exist. The *robot types model*, which considers sets of robots with identical capabilities, can be converted to the services model (Service & Adams, 2011a). The *resources model* considers tasks' resource (i.e., sensor and actuator) requirements and if resources are collocated or provided by different robots (Vig & Adams, 2007). Certain problems (e.g., identifying the lowest cost coalition for a task) are more computationally difficult than with the services model (Service & Adams, 2011a), and more knowledge of robot hardware is needed for task design. The *schema model* determines the information flow from robots' environmental sensors to code modules to actuators (Tang & Parker, 2005). This model allows for very flexible task performance, but considers only motion-based (i.e., not processing or sensing) tasks. This manuscript focuses on applying the services model.

## 3.2 Collective coalition formation

Coalition formation algorithms for robotic collectives, as opposed to smaller multiple robot systems, typically do not solve coalition formation as defined in Sect. 1.1. Instead, the algorithms solve biologically-inspired, reduced difficulty problems that consider *inexact team sizes*, which can be a fraction of the collective or depend on mission conditions (Huang & Robinson, 1992; Theraulaz et al., 1998). The resulting problem formulations generally do not optimize mission utility. Recently, *hedonic games* have also been applied to collective coalition formation to solve problems more similar to Sect. 1.1, optimizing utility and specifying target team sizes. Approaches with inexact team sizes and leveraging hedonic games are surveyed.

### 3.2.1 Inexact team sizes

Many inexact team size coalition formation approaches model biological collectives' behaviors. *Stimulus–response* is based on entomology (Theraulaz et al., 1998) and sociology (Granovetter, 1978), in which each individual chooses from a set of mutually exclusive options based on a combination of individual preference and external stimuli. For example, individuals may or may not participate in a continuous resource-gathering task based on how close the collective is to its desired resource level (Castello et al., 2014; Krieger et al., 2000; Rosenfeld et al., 2006; Yongming et al., 2010) or the amount of interference between robots (Pang et al., 2019). Stimulus–response algorithms have also been used for continuous cooperative monitoring tasks (Low et al., 2004), and to distribute approximate numbers of robots to distinct task sites (Kanakia et al., 2016). These algorithms can scale well,

as they require little to no communication, and the computational requirements depend on the various specified thresholds, not the collective size; however, these models are not directly transferable to the target collective coalition formation optimization problem, as there is not an existing mapping from coalition formation's characteristic function to stimulus–response models' thresholds and stimuli.

Algorithms that distribute approximate fractions of a collective among tasks have also incorporated the *activator/inhibitor model* that represents labor division in bee colonies (Huang & Robinson, 1992; Xiao & Wang, 2018). The algorithms' agents are modeled as having an internal activator that causes them to progress through a task sequence, while progression can be inhibited by factors, such as the number of robots per task (Zahadat et al., 2015; Zahadat & Schmickl, 2016). These algorithms require relatively simple computation and rely predominantly on distributed local communication; however, they do not solve the desired optimization problem, as they assume that all tasks will be assigned robots without considering task utility. Recall that optimizing task utility is important in the target domain, as it enables the collective to maximize its achievement of mission objectives.

Another notable algorithm family with inexact team sizes uses ordinary differential equations modeling the robots' transition rates between tasks (e.g., Berman et al., 2009b; Elamvazhuthi et al., 2018; Hsieh et al., 2008; Mather & Hsieh, 2011). Individual robot controllers are derived from these equations and determine robots' stochastic transitions between tasks. Equilibrium is reached when each task is assigned approximately the desired fraction of the collective. A key advantage of these algorithms is that the complexity of the robots' controllers does not increase with the collective size, facilitating scaling (Berman et al., 2009b). Some algorithms were evaluated with as many as 2,000 (Mather & Hsieh, 2011), and even 20,000 (Hsieh et al., 2008) simulated point-based robots. Additionally, algorithm performance is distributed and requires no communication between robots when visually estimating the number of robots currently assigned to each task is possible. Otherwise, the robots at a task site must communicate with each other. However, the majority of these algorithms consider only homogeneous collectives, and none consider the services model. A service extension will likely be difficult, as the number of transitions that must be considered increase substantially with the number of services and services per robot. These algorithms also do not transfer directly to the coalition formation optimization problem, as they assign fractions of the collective to tasks without addressing task utilities. Finally, a key limitation is that robots travel between task sites during coalition formation. Quadcopters can potentially travel quickly, but doing so wastes their already very limited battery lives (Diehl & Adams, 2021b). Conversely, ground robots generally have longer battery lives, but may require substantial time to travel between tasks, as the area of operation for the target applications can be large and highly deconstructed; thus, this family of algorithms is excluded from the empirical evaluation.

### 3.2.2 Hedonic games

Recent collective coalition formation leverages *anonymous hedonic games*, where coalition size determines utility, and each robot tries to optimize its individual utility by joining and leaving coalitions until *Nash stability* is reached (i.e., robots cannot benefit by individually changing coalitions) (Drèze & Greenberg, 1980). Centralized hedonic games (Czarnecki & Dutta, 2019, 2021) are not well-suited to collectives, as they require frequent communication with a centralized computer (Sect. 2.4). Distributed algorithms are most relevant.

The distributed algorithms (Jang et al., 2019, 2018; Dutta et al., 2021) derived from the GRoup Agent Partitioning and Placing Event (GRAPE) algorithm (Jang et al., 2018), in which each robot joins its preferred coalition, sends the resulting coalition structure to its network neighbors, and updates its coalition structure based on neighbors' messages. The GRAPE algorithm variations allow for homogeneous systems (Jang et al., 2018), variations in the performance of a single service (Jang et al., 2019), or heterogeneity under the resources model (Dutta et al., 2021); however, no services model variant exists.

The GRAPE variations are fast (i.e., milliseconds with 50 robots) and use local communication (Jang et al., 2019, 2018; Dutta et al., 2021); however, evaluation with large, heterogeneous collectives is an open problem. Additionally, individual robot utilities are not equivalent to coalition utilities, and differences may create suboptimal solutions (Jang et al., 2019, 2018).

### 3.3 Multiple robot coalition formation

Multiple robot coalition formation or *ST-MR task allocation* (i.e., a robot performs a Single Task and a task is performed by Multiple Robots) is well-studied (Gerkey & Mataric, 2004; Korsah et al., 2013). The pros and cons of the primary algorithms are outlined.

#### 3.3.1 Exact and approximation algorithms

Exact algorithms return optimal solutions, while approximation algorithms return solutions within a known factor of optimal. These solution quality guarantees can be advantageous; however, exact and approximation algorithms have not been favored historically for multiple robot coalition formation, due to coalition formation's high computational complexity (Service & Adams, 2011a; Zhang & Parker, 2013).

These algorithm types are more feasible if the search space is constrained. Dynamic programming with a limited number of robot types,  $j < O(n)$ , produces optimal solutions with a  $O(n^{2j}m)$  computational complexity (Service & Adams, 2011a), compared to the  $O(3^n)$  otherwise required (Rothkopf et al., 1998). Limited robot types (see Sect. 2.3) can be a valid assumption; however, the dynamic programming algorithm is centralized, which is less compatible with collectives (see Sect. 2.4).

Adaptations of Shehory and Kraus's distributed software multi-agent algorithm incorporate a maximum coalition size constraint,  $k < O(n)$  (Shehory & Kraus, 1995, 1998). The algorithms have two stages (Vig & Adams, 2007, 2006b). First, each robot calculates its possible coalitions' values. Next, robots broadcast their best coalitions. The best overall coalition is greedily added to the coalition structure until all tasks have been assigned or no coalitions can perform the remaining tasks. The base computational complexity is  $O(n^k m)$ , and the approximation ratio depends on the heuristic used to identify the best coalition (Shehory & Kraus, 1995, 1998).

There are three common heuristics. *Maximum utility* selects the highest utility coalition and finds solutions within a factor  $k + 1$  of the optimal utility (Service & Adams, 2011a). *Minimum cost* selects the lowest cost (i.e.,  $1/\text{utility}$ ) coalition and finds solutions within a factor  $O(k/\log k)$  of the optimal cost (Shehory & Kraus, 1998). *Average utility* finds the coalition with the highest utility to size ratio, producing solutions within  $2k$  of the optimal utility (Zhang & Parker, 2013). This heuristic is useful when smaller coalitions are favored, and the utility does not depend on coalition size.



The *resource-centric heuristic* aims to prevent robots that are needed for many tasks from being assigned early (Zhang & Parker, 2013). The heuristic ranks coalitions based on utility and the degree to which their selection will limit future task allocation. The approximation factor is  $(2k + 2)$ , and the computational complexity is  $O(\min(n, m)m^2 \binom{n}{k}^2)$  (Zhang & Parker, 2013).

Any algorithm variation can incorporate the Fault Tolerance Coefficient to balance assigning redundant robots, in case some fail to perform tasks, with the cost of using extra robots (Vig & Adams, 2006b). Additionally, other algorithm variations exist. Service and Adams incorporated the services model and bipartite matching, reducing the computational complexity to  $O(n^{3/2}m)$  (Service & Adams, 2011a). Dutta et al.'s one-to-many bipartite matching-based algorithm has a  $O(nm)$  computational complexity and a  $(1/(1 + \max(k - \Delta)))$  cost approximation ratio, where  $\Delta \in \{0, 1\}$ ; however, it only allows for homogeneous systems (Dutta & Asaithambi, 2019).

The advantage of the greedy approximation algorithms is that they provide solution quality guarantees, are flexible in the objective functions that they can optimize, and facilitate heterogeneity. However, their reasonable computational complexities rely on a small maximum coalition size,  $k$ , which is unlikely given collectives' scale. The algorithms also rely on all-to-all broadcasts to communicate coalition values among the robots, which may be infeasible with deployed disaster response and military networks. Thus, existing variations cannot be easily applied to collective coalition formation.

### 3.3.2 Auction algorithms

Auctions are a market-based resource allocation method, where buyers and sellers exchange information about the price at which they are willing to buy and sell goods (Phelps et al., 2008). A seller aims to obtain a high price, while a buyer's goal is to obtain goods for a low price. Auction algorithms are popular for multiple robot coalition formation, because they are an intuitive method of distributing resources. Existing algorithms vary based on the method of mapping tasks and robots to buyers and sellers, as well as the auction type.

Some algorithms have robots bid directly on tasks. First-price, one round auction-based algorithms auction off tasks in the order that they are received (Gerkey & Matarić, 2002; Sujit et al., 2008). An auctioneer broadcasts a task's required resources to the robots, robots reply with the cost of their participation, and the auctioneer chooses the lowest cost coalition that can perform the task. The advantage is fast runtimes relative to other auction mechanisms; however, a less valuable task received before a more valuable task is given precedence. Double auctions address this limitation by auctioning off all tasks at once (Guerrero et al., 2017; Xie et al., 2018). Robots in Guerrero et al.'s double auction algorithm send task auctioneers information about their relevant abilities, and task auctioneers send back bids based on the utility of the tasks they represent (Guerrero et al., 2017), while in Xie et al.'s algorithm, the tasks start the auction process (Xie et al., 2018). These algorithms allow tasks to compete for resources, at the cost of additional communication.

Robots can also bid on tasks through intermediaries. Project Manager-Oriented Coalition Formation has robots elect project managers for each task, which select coalitions of their neighbors in the network topology (Oh et al., 2017). This approach ensures that robots in a coalition can communicate, but severely restricts the solution space. The Automated Synthesis of Multi-robot Task solutions through software Reconfiguration (ASyMTRe) algorithm variants have robots form coalitions that bid on tasks

(Zhang & Parker, 2012; Tang & Parker, 2007). All ASyMTRe variations use the schema model. Finally, other algorithms have tasks bid on robot capabilities. The Robot Allocation through Coalitions Using Heterogeneous Non-Cooperative Agents (RACHNA) approach uses an ascending auction, with tasks bidding on services via service agents (Vig & Adams, 2006a). A limitation is that robots can be unnecessarily reassigned. A simultaneous descending auction addresses this limitation, at the cost of an empirically higher runtime (Service et al., 2014).

Overall, auction-based approaches are more applicable to collectives than exact and approximation algorithms, as they do not make assumptions about coalition size or require communication with a central computer. However, the requirement that all robots communicate with the task auctioneers (or task agents with service agents) may be difficult for military and disaster response networks (Vig & Adams, 2006a; Service et al., 2014). The communication required is substantial and can increase at least linearly with the collective size, potentially exceeding deployed network's limited bandwidth. Additionally, algorithms were generally not evaluated with large collectives, so it is necessary to assess scalability.

### 3.3.3 Biologically-inspired algorithms

Early scalable robot coalition formation leveraged biologically-inspired optimization mechanisms. The simulated Annealing inspired ANT colony optimization (sA-ANT) algorithm, which leverages ant colony optimization and simulated annealing, allocates only one task at a time (Sen & Adams, 2013b). Double-layered ant colony optimization addresses this limitation (Yeh & Sugawara, 2016). Additionally, particle swarm optimization and the Pareto Archived Evolution Strategy-based algorithm addressed multi-objective coalition formation (Agarwal et al., 2014; Mouradian et al., 2017).

Most algorithms were evaluated in simulation at or near the scale of collectives (Agarwal et al., 2014; Mouradian et al., 2017; Sen & Adams, 2013b); however, the algorithms are centralized and either allocate a single task at a time (Mouradian et al., 2017; Sen & Adams, 2013b) or their experimental run times indicate requiring over 15 min for even small collectives (Agarwal et al., 2014; Yeh & Sugawara, 2016).

Haque et al. proposed a decentralized algorithm based on alliance formation between male bottlenose dolphins (Haque et al., 2013); however, this algorithm restricts coalition sizes, which is not well suited to collectives.

## 3.4 Summary

The algorithm types most likely to be suitable for collectives, given the coalition formation optimization problem, are hedonic games and auctions. Hedonic games produce solutions quickly with multiple robot systems and can be distributed; however, distributed hedonic games have not been evaluated with large heterogeneous collectives (e.g., 1000 robots), and communication requirements have generally not been evaluated. Auctions are promising, given that some existing algorithms (e.g., Zhang & Parker, 2012; Xie et al., 2018; Vig & Adams, 2006a) are decentralized and consider heterogeneous systems; however, auctions also have not been evaluated at the scale of collectives and may require excessive communication.

## 4 Algorithms

Hedonic game and auction coalition formation algorithms were evaluated for the viability of their use with collectives. Homogeneous GRAPE, a distributed hedonic game-based algorithm, is evaluated, as there is no services model variant (Jang et al., 2018). The distributed auction-based algorithms are RACHNA (Vig & Adams, 2007, 2006a) and simultaneous descending auction (Service et al., 2014), which incorporate common auction protocols.

### 4.1 GRAPE

GRAPE incorporates an anonymous hedonic game, meaning that each robot joins the most individually profitable coalition, as determined based on coalition size, not the coalition members' identities (Jang et al., 2018). During each iteration of GRAPE, robots select their highest valued coalition, broadcast their beliefs, about all robots' current task assignments, and update their belief states based on messages from neighboring robots in the network topology. A robot's message is given precedence if the robot's belief state has been updated more times, or the same number of times and more recently, than the receiving robot's. This precedence system serves as a *distributed mutex* that allows only a single robot to alter the valid coalition assignments during each iteration. The algorithm completes when a Nash stable partition is derived.

GRAPE requires  $O(m)$  computation and  $O(n)$  communication per iteration on a single robot, where  $m$  is the number of tasks, and  $n$  is the collective size (Jang et al., 2018). A Nash stable partition can be found in  $O(n^2)$  iterations, given a fully connected network topology, resulting in  $O(n^2m)$  computational and  $O(n^3)$  communication complexities (Jang et al., 2018). Any connected network topology with a diameter  $d_G$  has a  $O(n^2md_G)$  computational complexity with a  $O(n^3d_G)$  communication complexity (Jang et al., 2018).

These complexities require that each robot's reward for any given task decreases as coalition size increases (Jang et al., 2018). This manuscript uses a *peaked reward*, or a system reward that is highest when a task is assigned a coalition with exactly the desired number of robots, divided evenly among coalition members (Jang et al., 2018). An individual robot's reward for task  $t_j$  with utility  $u_j$  that requires  $n_j$  robots and is assigned to coalition  $C_j$  is:

$$utility(t_j, C_j) = \frac{u_j}{n_j} \times e^{-\frac{|C_j|}{n_j} + 1}, \quad (1)$$

If multiple service types are available, this function cannot determine which robots will perform each service type; thus, GRAPE is considered only for homogeneous collectives with one service type.

### 4.2 RACHNA

RACHNA incorporates a *combinatorial ascending auction*, meaning that buyers bid on bundles of goods, and bids increase as the auction progresses (Vig & Adams, 2006a). The sellers are *service agents* (i.e., one agent per service type), and the buyers are *task agents*. Service agents sell a service type and track robots' current salaries (i.e., rewards for coalition membership) (Vig & Adams, 2006a). Task agents bid on service bundles, which they

are awarded if the bid is at least the robots' salaries, plus  $|C_j| \times \epsilon_{inc}$ , where  $C_j$  is the task's coalition and  $\epsilon_{inc}$  is a fixed wage increase. A task's maximum bid is its utility.

At most  $n(u_{max}/\epsilon_{inc})$  bidding rounds occur, where  $u_{max}$  is the highest task utility, with  $O(m)$  bids per round. Task agents require  $O(n)$  computation to determine their bids, and the  $|S|$  service agents require  $O(n \log n)$  computation to process bids, resulting in  $O(mn^2 \log n(u_{max}/\epsilon_{inc}))$  total computation. A service agent communicates with at most all task agents and robots per round (Vig & Adams, 2006a). Messages exchanged with task agents are size  $O(n)$ , and messages exchanged with the robots have size  $O(1)$ ; thus, the communication complexity is  $O(nm|S|)$  per round and  $O(n^2 m|S|(u_{max}/\epsilon_{inc}))$  in total.

Prior evaluation of RACHNA treated  $\epsilon_{inc}$  as a fixed parameter controlling the degree of competition (Vig & Adams, 2007, 2006a; Sen & Adams, 2013a). However, tasks for which  $|C_j| \times \epsilon_{inc} > u_j$  cannot be assigned coalitions, even when sufficient robots exist. This limitation is incompatible with collectives' potentially large coalition sizes; thus, RACHNA must incorporate collective size into its  $\epsilon_{inc}$  threshold dynamically.  $RACHNA_{dt}$  denotes a new implementation of RACHNA where  $\epsilon_{inc} = 1/n$ .  $RACHNA_{dt}$ 's computational complexity is  $O(mn^3 \log n(u_{max}))$ , and the communication complexity is  $O(n^3 m|S|(u_{max}))$ .

### 4.3 Simultaneous descending auction

Simultaneous descending auction, like RACHNA, incorporates a combinatorial auction with service and task agents. Each robot's salary is set initially to the maximum task utility, plus  $\epsilon_{dec}$ , where  $\epsilon_{dec}$  is a fixed wage decrement. Robot's salaries are decremented by  $\epsilon_{dec}$  at the beginning of each bidding round, and task agents that still require additional services bid on service bundles. If the task's utility is higher than the total salaries of all robots in the bundle, the task agent is able to afford a service bundle. The auction stops when all robots have been purchased or all salaries are zero.

Two implementations are considered. The first, simultaneous descending auction with a small coalition optimization (denoted  $SDA_{SCO}$ ), has task agents determine their bids by enumerating all coalitions that meet their service requirements. The computational complexity is  $O(mn^k)$  per iteration, where  $k$  is the maximum coalition size (Service et al., 2014). There are at most  $u_{max}/\epsilon_{dec}$  iterations, resulting in an overall computational complexity of  $O(mn^k u_{max}/\epsilon_{dec})$ . This implementation is faster than the alternative with small coalitions,  $k \leq 3$ , and was evaluated in prior work (Service et al., 2014).

The second implementation, denoted  $SDA_M$ , determines task agents' bids using weighted bipartite matching (Service et al., 2014). The computational complexity is  $O(mn^4 u_{max}/\epsilon_{dec})$ , regardless of coalition size (Service et al., 2014). This complexity is identical to  $SDA_{SCO}$ 's when  $k = 4$  and faster when  $k > 4$  (Service et al., 2014). Both implementations have communication complexities of  $O((m+n)|S|u_{max}/\epsilon_{dec})$ , as service agents communicate with at most all robots and task agents during each iteration. All considered algorithms' reward/cost functions are summarized in Table 2.

## 5 Experimental design

A simulation-based experiment assessed each algorithm's *viability*, where a viable algorithm produces near-optimal solutions in near real-time for a range of collective coalition formation problems, as required for highly dynamic domains. The target thresholds are  $< 5$  min runtimes and  $> 95\%$  utility solutions. Relaxed thresholds of  $< 10$  min runtimes and

**Table 2** Reward/cost function by algorithm

Algorithm	Type	Function
GRAPE	Reward	$\frac{u_j}{n_j} \times e^{-\frac{ C_j }{n_j} + 1}$
RACHNA	Cost	$(\sum_{a_i \in C_j} cost(a_i)) + (\epsilon_{inc} \times  C_j ), \epsilon_{inc} = 1$
RACHNA <sub>dt</sub>	Cost	$(\sum_{a_i \in C_j} cost(a_i)) + (\frac{1}{n} \times  C_j )$
SDA <sub>SCO</sub>	Cost	$\sum_{a_i \in C_j} cost(a_i), \epsilon_{dec} = 1$
SDA <sub>M</sub>	Cost	$\sum_{a_i \in C_j} cost(a_i), \epsilon_{dec} = 1$

> 90% utility solutions are also considered to account for the fact viability is application-dependent, and some deployments may have more lenient requirements.

Recall that communication resources in the target domains are often very limited and easily overwhelmed (Klinsompus & Nupairoj, 2015). For example, cellular communication was down for several days during the 2023 Maui wildfires, leaving responders with only minimal, deployable communication networks (Kelly, 2023). There is no standard threshold for acceptable communication limits, as limits can depend on a number of factors (e.g., routing protocols, communication medium, network topology, developments in networking technology, the quantity of communication required for task performance); however, the objective in all instances is to use minimal communication, preserving resources for the transmission of data relevant to task performance. This manuscript assesses communication by comparing the algorithms. The total communication target is < 500 MB, corresponding to the lowest maximum communication empirically required for the majority of the considered coalition formation problems. An algorithm that satisfies this target requires minimal communication in that it requires no more communication than the overall lowest-communication algorithm. A relaxed < 600 MB total communication target was also considered.

The experiment considered *achievable missions*, meaning that the collectives' robots offered sufficient services to perform all tasks simultaneously. Real-world collective deployments will ideally incorporate achievable missions but are unlikely to involve substantially more than the minimum required robots, due to the expense and logistical challenges of deploying large collectives. The experiment's collectives possessed exactly the minimum number of robots required (i.e., coalition formation solutions must utilize all robots).

The independent variables were the algorithm, collective size, number of tasks, and collective composition (see Table 3). The algorithms considered were GRAPE, RACHNA

**Table 3** Independent variables

Algorithms	GRAPE, RACHNA, RACHNA <sub>dt</sub> , SA <sub>SCO</sub> and SA <sub>M</sub>
Collective Size	25, 50, 100, 500, 1000
Percent Tasks	1%, 10%, 50% (i.e., coalition sizes 100, 10, 2)
Service Types	1, 5, 10
Services Per Robot	1, 5

( $\epsilon_{inc} = 1$ ),  $RACHNA_{dr}$ ,  $SDA_{SCO}$  ( $\epsilon_{dec} = 1$ ), and  $SDA_M$  ( $\epsilon_{dec} = 1$ ). The *collective size* varied from multiple robot systems (i.e., 25–50 robots) to collectives (i.e., 100–1000 robots). The numbers of tasks were 1%, 10%, and 50% the number of robots, corresponding to average coalition sizes 100, 10, and 2, respectively. 1% tasks corresponds to large-scale tasks (e.g., establishing network coverage or searching for survivors), while 10% tasks corresponds to medium-scale tasks (e.g., assessing a damaged building or creating a cordon). Examples of small-scale 50% tasks include guarding a building entrance or providing a medical evacuation. 1% tasks (i.e., coalition size 100) was used only for collectives with  $> 100$  robots, as a mission is only achievable if the number of robots per task is smaller than the collective size. Collective composition encompassed the number of *service types* and *services per robot*, where more capable robots had more services, and the number of service types and services per robot combinations determined the level of heterogeneity. A collective was homogeneous if the numbers of service types and services per robot were equal. Otherwise, the collective was heterogeneous.

GRAPE was considered only for homogeneous collectives with one service type, as GRAPE does not incorporate a services model. Additionally,  $SDA_{SCO}$  was considered only for problems with multiple robot systems (i.e., 25 or 50 robots), due to the algorithm's high computational complexity. The other algorithms were considered for all independent variable value

The experiment used a centralized C++ simulator on a HP Z640 Workstation (Intel Xeon processor, 62 GB RAM) (Sen & Adams, 2013a). The simulator performed each algorithm iteration for each robot sequentially and assumed a fully connected communication topology. The robots' embodiment was partially represented by the services model; however, the robots' positions were not considered, as incorporating positions in the utility function, when applicable, is highly dependent on the application environment and robot hardware. Twenty-five problem instances were randomly generated per independent variable combination, where a problem instance comprised sets of robots and tasks, each with associated services. Robots' and tasks' services were selected randomly, and each task was assigned a random integer utility in the range [1, 50]. Each trial, or problem, was allocated twelve hours to produce a solution. This time limit is too long for high tempo dynamic domains (e.g., disaster response) or even short term pre-mission planning (e.g., 2–4 h breaks between mission deployments (Defense Advanced Research Projects Agency, 2019)). All algorithms can produce partial solutions, but this capability was not evaluated.

The dependent variable *success rate* represents the ratio of problems for which algorithms provided non-zero utility solutions within the time limit. A trial was *unsuccessful* if the computer's memory limit was exceeded, the algorithm's runtime exceeded the 12-hour time limit, or the algorithm was otherwise unable to assign any robots to tasks. Only successful trials were considered when analyzing the other dependent variables, as unsuccessful trials can cause less suitable algorithms to appear to perform well under certain metrics. For example, an algorithm that quickly exceeds the memory limit, resulting in no viable solution, will have a very low runtime, but is less suitable than an algorithm that produces a solution given a longer runtime.

The other dependent variables are runtime, total communication, and percent utility. *Runtime* is the time in minutes (min) and seconds (s) required for an algorithm to produce a solution. *Total communication* is the sum of all message sizes in megabytes (MB). *Percent utility* measures the solution quality of successful trials (i.e., solution utility/optimal utility). Overall, higher success rates and percent utilities with lower runtimes and total communication are preferred. It was *hypothesized* that none of the algorithms will perform well for all metrics across the range of collective sizes and compositions.

## 6 Results

Results are presented for homogeneous and heterogeneous collectives. Box plots were used, as the data was not normally distributed. Non-parametric statistical methods assessed significance across each independent variable. Mann–Whitney–Wilcoxon tests compared the numbers of service types and the services per robot, while Kruskal–Wallis analysis with Mann–Whitney–Wilcoxon post-hoc tests compared across the collective sizes and percent tasks. All analysis included only independent variable combinations for which an algorithm had successful trials across all independent variable values (e.g., percent task analysis considered only 500 and 1000 robot collectives). An algorithm was deemed viable if it had consistently low runtimes and communication, as well as high percent utilities, for all independent variable values.

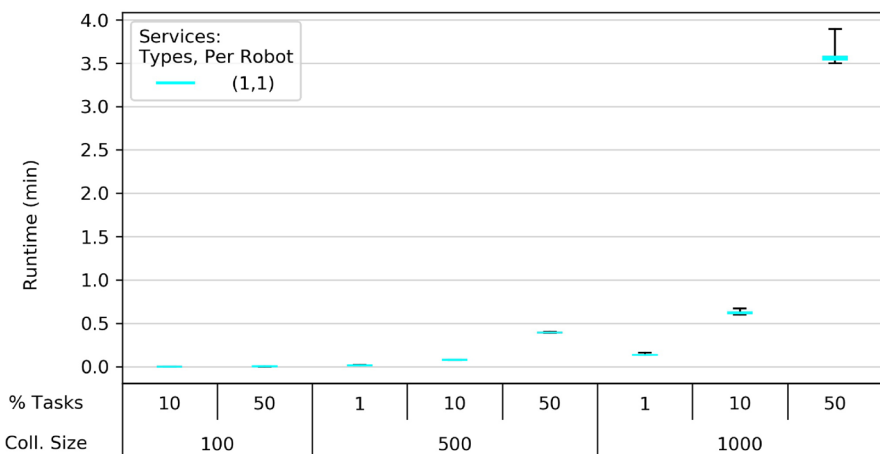
### 6.1 Homogeneous collective results

A total of 600 trials with homogeneous collectives were run for each of RACHNA, RACHNA<sub>dt</sub>, and SDA<sub>M</sub>. Recall that GRAPE's analysis considered only the 300 trials with one service type, and SDA<sub>SCO</sub> was considered for only the 200 multiple robot trials.

#### 6.1.1 GRAPE

GRAPE produced optimal solutions for all trials with one service type, for a 50% overall success rate. Recall that GRAPE was unable to perform the 300 trials with five service types, or half the trials.

All successful trials had low runtimes, well within the < 5 min target (Fig. 1). Runtimes did increase substantially from multiple robot systems (< 1 s) to collectives (< 3 min 53 s), as well as with percent tasks. The increases were significant with collective size ( $H(n = 50) = 238.94, p < 0.01$ ) and percent tasks ( $H(n = 50) = 74.50, p < 0.01$ ),



**Fig. 1** GRAPE's runtimes (min) with homogeneous collectives. GRAPE did not solve problems with 5 service types and 5 services per robot. All multiple robot trials (i.e., 25–50 robots) completed in < 1 s. Note that the y-axis maximum is 4 min

with significant post-hoc analyses ( $p < 0.01$ ). The longest runtimes were for 1000 robot collectives with 50% tasks (i.e., the largest number of robots and tasks), consistent with GRAPE's  $O(n^2m)$  computational complexity. Nevertheless, all of GRAPE's runtimes were sufficiently fast for high-tempo applications.

GRAPE's total communication increased only with collective size (Table 4). This increase was significant ( $H(n = 50) = 249.00, p < 0.01$ ), with significant post-hoc analyses ( $p < 0.01$ ). Recall that GRAPE's per iteration communication complexity is  $O(n)$ . Percent tasks and problem instance did not impact the communication requirement, because all problems required  $n$  iterations (i.e., no robots deviated from their initially selected coalitions).

### 6.1.2 RACHNA and RACHNA<sub>dt</sub>

RACHNA<sub>dt</sub> solved all problems, while RACHNA produced solutions for only 82.3% of trials. All RACHNA trials with 1% tasks were unsuccessful, as well as 12% with 25 robots, 10% tasks, and either collective composition. All unsuccessful trials considered problems in which every tasks' required coalition size exceeded its utility, consistent with RACHNA's known limitation.

RACHNA's and RACHNA<sub>dt</sub>'s runtimes differed by  $< 1$  min (see Fig. 2). Both algorithms had low runtimes with multiple robot systems ( $< 1$  s), which increased substantially for collectives ( $< 125$  min 48 s). Runtimes also increased with percent tasks, consistent with the  $O(mn^3 \log n)$  per iteration computational complexity. Additionally, runtimes increased with the number of service types, which corresponds to an increased number of service agents performing computation. The longest runtimes were for 1000 robots, 50% tasks, and five service types. These runtimes were much too long for high-temp domains.

RACHNA's runtime increase with respect to collective size was significant ( $H(n_{25} = 44, n = 50) = 458.75, p < 0.01$ ), with significant post-hoc tests ( $p < 0.01$ ). Note that the lower number of samples for a collective size of 25 is due to unsuccessful trials. RACHNA's runtime also increased significantly when the percent tasks increased from 10% to 50% ( $p < 0.01$ ) and the service types increased from one to five ( $p = 0.01$ ).

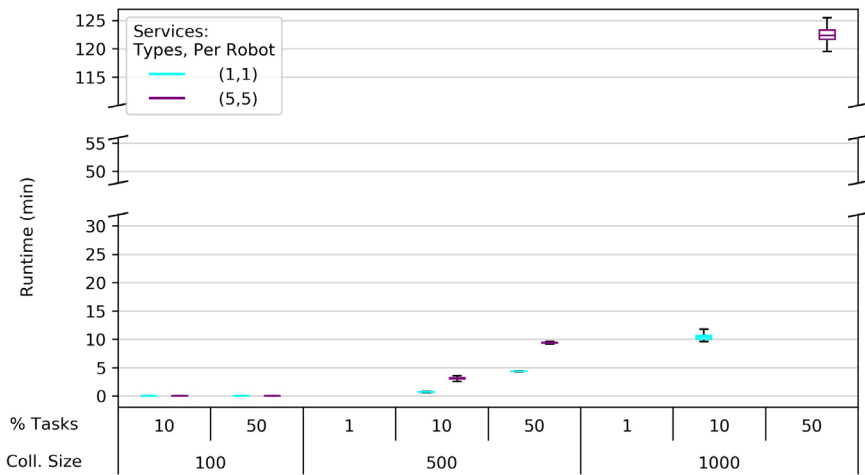
RACHNA<sub>dt</sub> produced similar results. RACHNA<sub>dt</sub>'s runtimes increased significantly with collective size ( $H(n = 50) = 467.32, p < 0.01$ ) and percent tasks ( $H(n = 50) = 138.08, p < 0.01$ ). Post-hoc tests found that all differences were significant ( $p < 0.01$ ). A significant difference also existed for increased numbers of service types ( $p < 0.01$ ).

RACHNA's and RACHNA<sub>dt</sub>'s communication results also had similar trends (Fig. 3), although RACHNA<sub>dt</sub> required up to an additional 600 MB. Both algorithms' communication increased substantially from multiple robot systems ( $\leq 70.81$  MB) to collectives

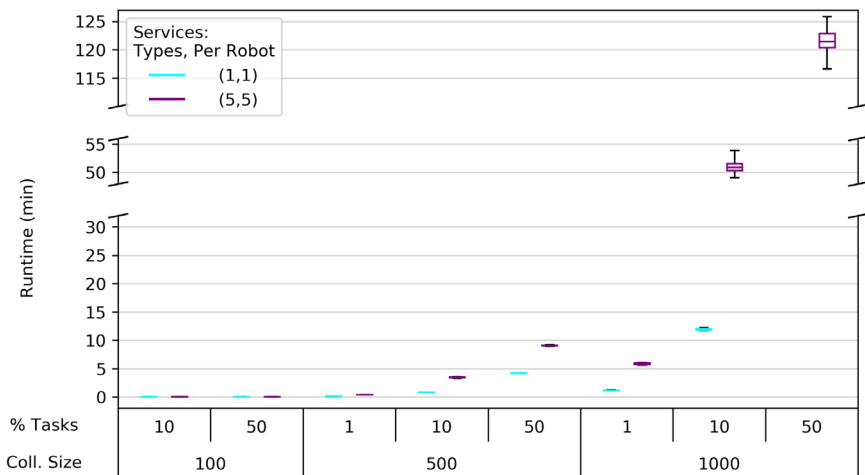
**Table 4** GRAPE's communication (MB) for homogeneous collectives with one service type. Communication was constant across problem instances with equal collective sizes and independent of other variables

Coll. Size	Communication (MB)
25	1.87
50	7.5
100	30
500	750
1,000	3,000





(a) RACHNA's runtimes with homogeneous collectives.

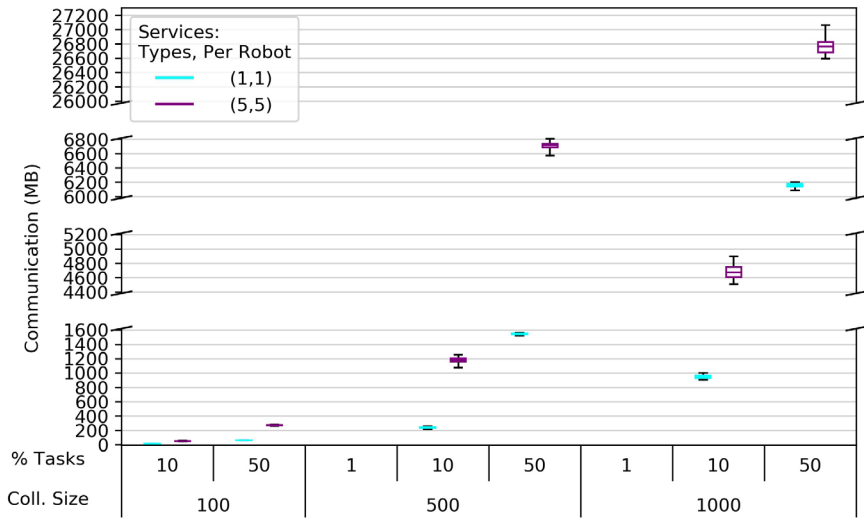


(b) RACHNA<sub>dt</sub>'s runtimes with homogeneous collectives.

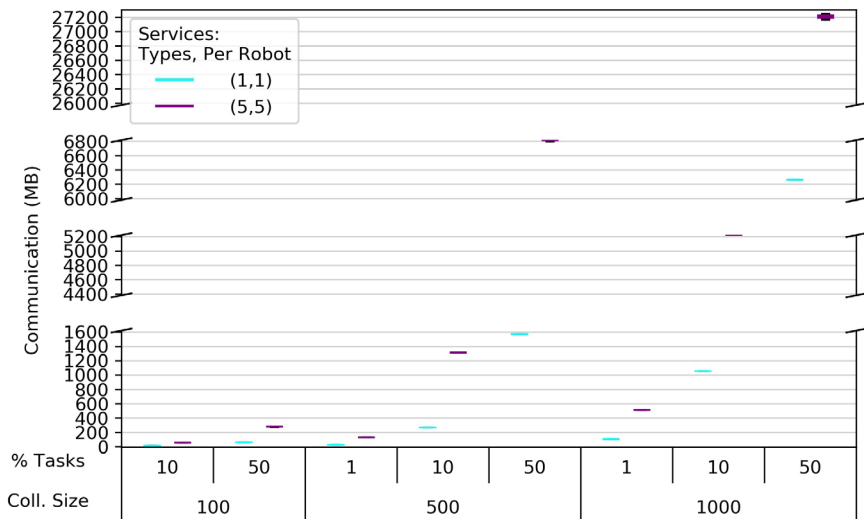
**Fig. 2** RACHNA's and RACHNA<sub>dt</sub>'s runtimes (min) with homogeneous collectives. All multiple robot trials (i.e., 25–50 robots) completed in < 1 s. RACHNA did not solve problems with 1% tasks. The y-axis maximum is 125 min (i.e., 2 h 5 min), and there are axis breaks

(≤27.25 GB), as well as with percent tasks and service types. 1,000 robot collectives with 50% tasks and five service types required the most communication, consistent with the algorithms  $O(nm|S|)$  per iteration communication complexity.

RACHNA's communication increase across collective sizes was significant ( $H(n_{25} = 44, n = 50) = 384.03, p < 0.01$ ), with significant post-hoc analyses ( $p < 0.01$ ). RACHNA's communication also increased significantly from 10% to 50% tasks ( $p < 0.01$ ) and one to five service types ( $p < 0.01$ ). Note that a Mann–Whitney–Wilcoxon test was



(a) RACHNA's communication with homogeneous collectives.

(b) RACHNA<sub>dt</sub>'s communication with homogeneous collectives.

**Fig. 3** RACHNA's and RACHNA<sub>dt</sub>'s communication (MB) with homogeneous collectives. All multiple robot trials (i.e., 25–50 robots) required < 70.81 MB. RACHNA solved no problems with 1% tasks. The y-axis maximum is 25,000 MB (25 GB)

used to assess significance with respect to percent tasks, as RACHNA had successful trials for only two values.

Similarly, RACHNA<sub>dt</sub>'s communication increased significantly with collective size ( $H(n = 50) = 390.97, p < 0.01$ ) and percent tasks ( $H(n = 50) = 234.59, p < 0.01$ ). Post-hoc tests identified significant differences between the collective sizes ( $p < 0.01$ ) and

percent tasks ( $p < 0.01$ ). A significant increase from one to five service types was also identified ( $p < 0.01$ ).

RACHNA produced near-optimal solutions for most problems (Table 5), but high variability resulted in low worst case utilities. Unassigned tasks were generally those requiring more robots than supported by the task utility, given RACHNA’s fixed threshold limitation. Such tasks occur more often with lower percent tasks (i.e., larger coalitions) and more substantially impact mission performance with fewer robots (i.e., fewer tasks for a given value of % tasks). Thus, multiple robot systems with 10% tasks (i.e., the lowest percent utility with successful trials) had the most variable percent utilities.

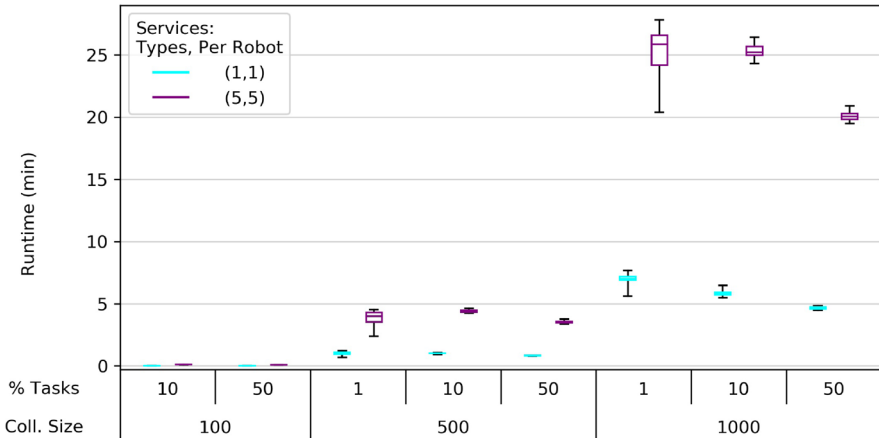
Significant differences existed across the collective sizes ( $H(n_{25} = 44, n = 50) = 53.16, p < 0.01$ ), specifically, between 25 and 100–1000 robots ( $p < 0.01$ ), 50 and 100–1000 robots ( $p = 0.02$  for collective size 100,  $p < 0.01$ , otherwise), and 100 and 1000 robots ( $p < 0.01$ ). No significant difference was found between 25 and 50 robots or 100 and 500 robots. 10% and 50% tasks also differed significantly ( $p < 0.01$ ), while one and five service types did not. RACHNA<sub>dt</sub> outperformed RACHNA, producing optimal solutions in all trials.

### 6.1.3 Simultaneous Descending Auction

The results for the  $SDA_M$  and  $SDA_{SCO}$  simultaneous descending auction implementations are presented separately, due to vastly different performance.  $SDA_M$  successfully produced optimal solutions in all trials.

**Table 5** RACHNA’s percent utility statistics with homogeneous collectives

Service Types	Percent Tasks	Collective Size	Percent Utilities Median (Minimum, Maximum)
1	10	25	100.0 (66.67, 100.0)
		50	99.17 (76.62, 100.0)
		100	96.77 (88.02, 100.0)
		500	96.39 (90.42, 99.040)
		1000	95.84 (92.94, 97.75)
	50	25	100.0 (98.91, 100.0)
		50	100.0 (99.6, 100.0)
		100	99.89 (99.56, 100.0)
		500	99.89 (99.8, 99.97)
		1000	99.88 (99.76, 99.92)
5	10	25	100.0 (72.73, 100.0)
		50	96.82 (80.91, 100.0)
		100	97.66 (90.17, 100.0)
		500	96.61 (92.88, 98.79)
		1000	96.02 (94.02, 97.89)
	50	25	100.0 (98.08, 100.0)
		50	99.93 (99.34, 100.0)
		100	99.84 (99.52, 100.0)
		500	99.91 (99.74, 100.0)
		1000	99.89 (99.83, 99.95)



**Fig. 4**  $SDA_M$ 's runtimes (min) with homogeneous collectives. All multiple robot trials (i.e., 25–50 robots) completed in < 1 s. The y-axis maximum is 25 min

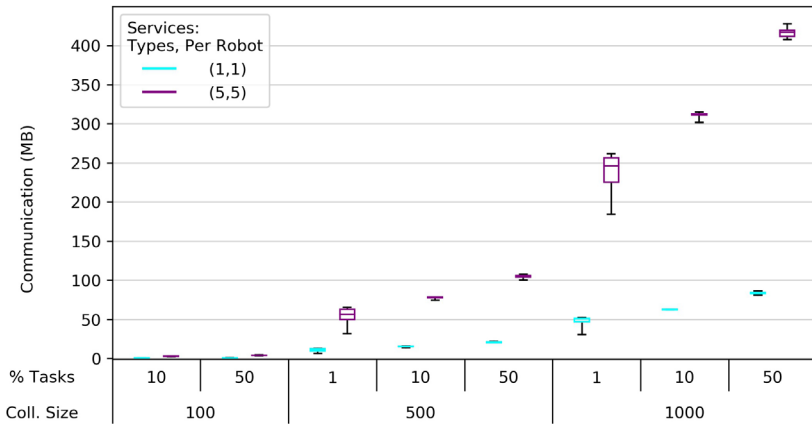
$SDA_M$ 's runtimes were within the < 5 min target for most trials, but increased substantially for collectives with five service types (Fig. 4). Multiple robot trials had < 1 s runtimes, which increased for collectives with one service type to  $\leq 6$  min 29 s. Collectives with five service types had runtimes  $\leq 26$  min 24 s. The increase with the number of service types can be attributed to the fact that task agents must consider each robot for more roles when determining their bids. Runtimes also decreased with the percent tasks, despite  $SDA_M$ 's  $O(mn^4)$  per iteration computational complexity. This trend can be explained by the fact that higher percent tasks correspond to smaller coalitions. The smaller coalitions are cheaper for the task agents, thus enabling tasks to bid in fewer iterations. The longest runtimes occurred with 1000 robots, five service types and 50% tasks.

The runtime increases across collective sizes ( $H(n = 50) = 465.37, p < 0.01$ ) and tasks ( $H(n = 50) = 10.61, p < 0.01$ ) were significant, as were the collective size post-hoc analyses ( $p < 0.01$ ); however, only differences between 10% and 50% tasks ( $p < 0.01$ ) and 1% and 50% tasks ( $p = 0.01$ ) were significant. No difference between 1% and 10% tasks was found. A significant increase from one to five service types was also identified ( $p < 0.01$ ).

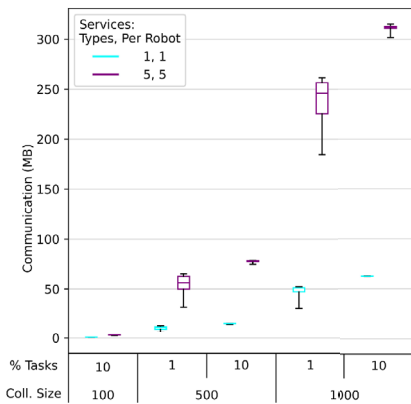
$SDA_M$ 's total communication was low with multiple robot systems (< 1.2 MB), but increased substantially for robotic collectives (see Fig. 5). Collectives with one service type used up to 86.29 MB total, while collectives with five service types used up to 428.03 MB. The total communication also increased with service types and percent tasks, consistent with the  $O((m+n)|S|)$  per iteration communication complexity. 1000 robot collectives with five service types and 50% tasks required the most communication.

Each of these trends was significant. Communication increased significantly across collective sizes ( $H$  ( $H(n = 50) = 435.46, p < 0.01$ ) and tasks ( $H$  ( $H(n = 50) = 35.17, p < 0.01$ ), with significant post-hoc analyses ( $p < 0.01$ ). A significant increase from one to five service types was also found ( $p < 0.01$ ).

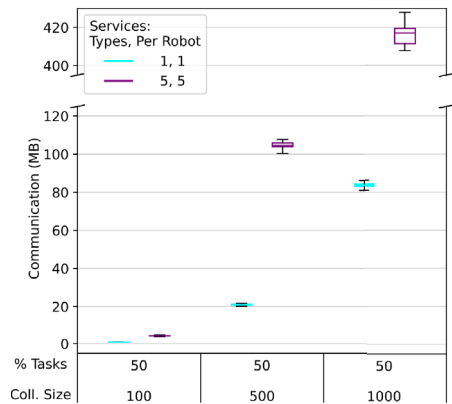
$SDA_M$ 's performance was generally better than  $SDA_{SCO}$ 's across all metrics. Recall that the  $SDA_{SCO}$  evaluation considered only the 200 problem instances with multiple robot systems (33.3% of all trials), due to  $SDA_{SCO}$ 's exponential computation requirement.  $SDA_{SCO}$ 's success rate with multiple robot systems was only 62.5%, for an overall success rate of 20.8%.



(a) Overview of  $SDA_M$ 's communication.



(b) Close-up with 1% and 10% tasks.



(c) Close-up with 50% tasks.

**Fig. 5**  $SDA_M$ 's communication (MB) with homogeneous collectives. All multiple robot trials (i.e., 25–50 robots) required < 1.2 MB. The y-axis varies between subfigures

The unsuccessful trials occurred when  $SDA_{SCO}$  exceeded either the computer's memory limit or the 12-hour runtime limit. The memory limit was exceeded for the 25 trials with 50 robots, 10% tasks, and one service type. The runtime limit was exceeded for the 50 trials with five service types and 10% tasks, regardless of the number of robots.  $SDA_{SCO}$  produced optimal solutions for the remaining 125 trials.

$SDA_{SCO}$ 's runtimes increased with the number of robots, service types, and decreased percent tasks, consistent with  $SDA_{SCO}$ 's  $O(mn^k)$  per iteration computational complexity (Table 6). Note that decreased service types decreases  $m$ , but also increases  $k$ . Each of the differences was statistically significant ( $p < 0.01$ ). Runtimes were also more variable.

The variation in runtimes was due to  $SDA_{SCO}$ 's sensitivity to differences in problem difficulty.  $SDA_{SCO}$ 's computational complexity is exponential in coalition size, and the longest runtimes for each independent variable combination were for the trials with the highest maximum coalition size. Variation in coalition size resulted in a worst case

**Table 6**  $SDA_{SCO}$ 's runtimes with homogeneous multiple robot systems. Collective trials (i.e., 100–1000 robots) were not attempted. No other trials were successful

Service Types	Percent Tasks	Collective Size	Runtime (min:s:ms) Median (Minimum, Maximum)
1	10	25	1:16:195 (0:08:040, 1:41:884)
	50	25	0:00:132 (0:00:72, 0:00:454)
		50	0:03:141 (0:00:769, 1:45:969)
5	50	25	0:01:918 (0:00:483, 8:41:294)
		50	1:4:711 (0:05:98, 400:34:373)

runtime (i.e., > 6 hours), much too long for high tempo applications or short term pre-mission planning.

The communication increased with the number of robots, tasks, and service types (Table 7), consistent with  $SDA_{SCO}$ 's  $O((m+n)|S|)$  communication complexity. Each of these increases was significant ( $p < 0.01$ ).

## 6.2 Homogeneous Collective Discussion

The homogeneous collective experiment assessed whether GRAPE, RAC-HNA, RACHNA<sub>dt</sub>,  $SDA_M$ , and  $SDA_{SCO}$  are suitable for very large homogeneous collectives in highly dynamic domains. The *hypothesis* that no algorithm fully satisfied the 100% success rate, < 5 min runtime, < 500 MB total communication, and > 95% utility criteria was supported. Algorithms also failed to meet the relaxed criteria (i.e., < 10 min runtime, < 600 MB communication, > 90% utility), denoted as the *reasonably close* criteria in Tables 8, 9 and 10.

$SDA_{SCO}$  satisfied the evaluation criteria for the fewest independent variable combinations (i.e., three out of twenty-four), as shown in Table 8. All criteria were satisfied for single-service systems with 10% tasks and 25 robots, as well as 50% tasks and 25–50 robots. Relaxing the target runtime to < 10 min enables  $SDA_{SCO}$  to satisfy all criteria for one additional independent variable combination (i.e., five services, 50% tasks, 25 robots). This relaxation is relatively reasonable for near real-time coalition formation; however, there are no other independent variable combinations for which  $SDA_{SCO}$  can satisfy all criteria, short of ignoring both the runtime and communication requirements entirely, which cannot be done in near real-time domains.

$SDA_{SCO}$ 's evaluation considered only multiple robot systems (i.e., ≤50 robots) due to a high computational complexity; however, it is clear from  $SDA_{SCO}$ 's poor performance

**Table 7**  $SDA_{SCO}$ 's communication with homogeneous multiple robot systems. Collective trials (i.e., 100–1000 robots) were not attempted. No other trials were successful

Service Types	Percent Tasks	Collective Size	Communication (MB) Median (Minimum, Maximum)
1	10	25	25.25 (2.92, 36.96)
		50	48.25 (31.39, 56.74)
		50	203.3 (165.96, 228.14)
5	50	25	239.83 (142.12, 280.6)
		50	1,001.37 (860.66, 1,109.07)

**Table 8** Homogeneous results summary:  $SDA_{SCO}$  and RACHNA. Each cell corresponds to a criterion for evaluating viability and an independent variable combination (i.e., an algorithm, services (S), percent tasks (T), and a collective size). A ✓ means that the algorithm met the criterion for all trials with the independent variable combination, while a ⚡ means that all trials were reasonably close to meeting the criterion. An ✗ means that the criterion was not met, or trials were not attempted due to known algorithm limitations. A - means that the independent variable combination is invalid

	S	T	$SDA_{SCO}$					RACHNA				
			Collective Size					Collective Size				
			25	50	100	500	1000	25	50	100	500	1000
Succ. Rate 100%	1	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✓	✗	✗	✗	✗	✗	✓	✓	✓	✓
		50	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
	5	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓
		50	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
Run-time < 5 min	1	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✓	✗	✗	✗	✗	✓	✓	✓	✓	⚡
		50	✓	✓	✗	✗	✗	✓	✓	✓	✓	✗
	5	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✗	✗	✗	✗	✗	✓	✓	✓	✓	✗
		50	⚡	✗	✗	✗	✗	✓	✓	✓	⚡	✗
Comm. < 500 MB	1	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✓	✗	✗	✗	✗	✓	✓	✓	✓	✗
		50	✓	✓	✗	✗	✗	✓	✓	✓	✗	✗
	5	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✗	✗	✗	✗	✗	✓	✓	✓	✗	✗
		50	✓	✗	✗	✗	✗	✓	✓	✓	✗	✗
Utility > 95%	1	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✓	✗	✗	✗	✗	✗	✗	✗	⚡	⚡
		50	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
	5	1	-	-	-	✗	✗	-	-	✗	✗	✗
		10	✗	✗	✗	✗	✗	✗	✗	⚡	⚡	⚡
		50	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓

with 25 and 50 robots that the algorithm will not scale to collectives.  $SDA_{SCO}$ 's major limitation is the excessive memory and computation required to enumerate the possible coalitions for each task. The evaluation criteria were satisfied only when there were few possible coalitions (i.e., 25 robots, 50% tasks, 1 service type), while performance degraded substantially with higher numbers of possible coalitions (i.e., increased service types, increased numbers of robots, decreased percent tasks). Homogeneous collectives are expected to possess large numbers of robots, as well as be assigned tasks that require large coalitions (e.g., assessing damage to a city block), which will result in even worse performance than occurred with multiple robot systems. Thus,  $SDA_{SCO}$  is not a viable approach for near real-time coalition formation with homogeneous collectives.

The next least viable algorithm was RACHNA (Table 8). RACHNA fully satisfied the evaluation criteria for only six independent variable combinations (i.e., one or five services with 50% tasks and 25–100 robots). Relaxing the runtime requirement to < 10 min and the utility requirement to > 90% enables RACHNA to satisfy all requirements for three additional independent variable combinations (i.e., one service with 10% tasks

**Table 9** Homogeneous results summary: RACHNA<sub>dt</sub> and GRAPE. Each cell corresponds to a criterion for evaluating viability and an independent variable combination (i.e., an algorithm, services (S), percent tasks (T), and a collective size). A ✓ means that the algorithm met the criterion for all trials with the independent variable combination, while a ⚡ means that all trials were reasonably close to meeting the criterion. An ✗ means that the criterion was not met, or trials were not attempted due to known algorithm limitations. A - means that the independent variable combination is invalid

	S	T	RACHNA <sub>dt</sub>					GRAPE				
			Collective Size					Collective Size				
			25	50	100	500	1000	25	50	100	500	1000
Succ. Rate 100%	1	1	-	-	-	✓	✓	-	-	-	✓	✓
		10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		50	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	5	1	-	-	-	✓	✓	-	-	-	✗	✗
		10	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗
		50	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗
Run-time < 5 min	1	1	-	-	-	✓	✓	-	-	-	✓	✓
		10	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓
		50	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
	5	1	-	-	-	✓	⚡	-	-	-	✗	✗
		10	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗
		50	✓	✓	✓	⚡	✗	✗	✗	✗	✗	✗
Comm. < 500 MB	1	1	-	-	-	✓	✓	-	-	-	✗	✗
		10	✓	✓	✓	✓	✗	✓	✓	✓	✗	✗
		50	✓	✓	✓	✗	✗	✓	✓	✓	✗	✗
	5	1	-	-	-	✓	⚡	-	-	-	✗	✗
		10	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗
		50	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗
Utility > 95%	1	1	-	-	-	✓	✓	-	-	-	✓	✓
		10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		50	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	5	1	-	-	-	✓	✓	-	-	-	✗	✗
		10	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗
		50	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗

and 500 robots, five services with 10% tasks and 100 robots, five services with 50% tasks and 500 robots). However, RACHNA cannot satisfy all evaluation criteria for the remaining independent variable combinations due to low percent utilities, as well as high worst-case runtimes and communication.

One of RACHNA’s limitations was its high runtimes and communication requirements with 500–1000 robots, which well exceeded the evaluation criteria. The other major limitation was RACHNA’s inability to assign coalitions to tasks with lower utilities than required coalition sizes. This limitation prevented RACHNA from solving any coalition formation problems with 1% tasks (i.e., 100 robot coalitions), as well as satisfying the utility criterion with 10% tasks (i.e., 10 robot coalitions). The limitation was masked somewhat for 10% tasks with increased collective size, due to the increased number of tasks. However, the percent utilities and success rates were lower than RACHNA<sub>dt</sub>’s, which, paired with comparable communication and runtimes, means that RACHNA<sub>dt</sub> is generally preferable to RACHNA.

RACHNA<sub>dt</sub> satisfied all evaluation criteria for sixteen independent variable combinations, including all combinations with multiple robot systems, as well as 100 robots (see



**Table 10** Homogeneous results summary:  $SDA_M$ . Each cell corresponds to a criterion for evaluating viability and an independent variable combination (i.e., services (S), percent tasks (T), and a collective size). A  $\checkmark$  means that  $SDA_M$  met the criterion for all trials with the independent variable combination, while a  $\checkmark$  means that all trials were reasonably close to meeting the criterion. An  $\times$  means that the criterion was not met. A - means that the independent variable combination is invalid

	S	T	$SDA_M$				
			Collective Size				
			25	50	100	500	1000
Success Rate 100%	1	1	-	-	-	$\checkmark$	$\checkmark$
		10	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
		50	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
	5	1	-	-	-	$\checkmark$	$\checkmark$
		10	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
		50	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Runtime < 5 min	1	1	-	-	-	$\checkmark$	$\checkmark$
		10	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
		50	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
	5	1	-	-	-	$\checkmark$	$\times$
		10	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$
		50	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$
Comm. < 500 MB	1	1	-	-	-	$\checkmark$	$\checkmark$
		10	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
		50	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
	5	1	-	-	-	$\checkmark$	$\checkmark$
		10	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
		50	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Utility > 95%	1	1	-	-	-	$\checkmark$	$\checkmark$
		10	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
		50	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
	5	1	-	-	-	$\checkmark$	$\checkmark$
		10	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
		50	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

Table 9). All criteria were also satisfied for single-service collectives with 500 robots and 1–10% tasks, as well as 1000 robots and 1% tasks, and five service collectives with 500 robots and 1% tasks. Relaxing the communication requirement to < 600 MB enables  $RACHNA_{dt}$  to satisfy the criteria for one additional independent variable combination (i.e., five services, 1% tasks, 100 robots).  $RACHNA_{dt}$ 's > 30 min runtimes and > 1 GB communication with all other independent variable combinations far exceeded the acceptable limits for near real-time domains.

RACHNA<sub>dt</sub>'s primary strength was that the success rate and utility criteria were always met, given sufficient time and communication. This performance makes RACHNA<sub>dt</sub> well-suited for applications in which pre-mission planning is possible (i.e., the tasks are known at least a few hours prior to deployment). However, RACHNA<sub>dt</sub>'s runtime and communication with 500–1000 robots are not suitable for near real-time domains, as the runtime and communication requirements are only met for some percent tasks (i.e., coalition sizes) and number of services. The number of services will generally be known prior to deployment, but the percent tasks will not, as tasks arise dynamically; thus, it generally cannot be guaranteed that RACHNA<sub>dt</sub> will be able to produce solutions within the runtime and communication requirements during any deployment with very large homogeneous collectives.

GRAPE fully satisfied the evaluation criteria for fewer independent variable combinations than RACHNA<sub>dt</sub> (i.e., the six combinations with one service and 25–100 robots), as shown in Table 9. GRAPE's primary limitation was its inability to solve any coalition formation problems with multiple service types, due to its lack of a services model. However, GRAPE with only one service type satisfied the success rate, runtime, and percent utility requirements for all independent variable combinations. The number of service types will generally be known prior to deployment, so GRAPE, unlike RACHNA<sub>dt</sub>, is suitable for deployment in near real-time domains, given single service collectives and sufficient communication infrastructure. Deployed networks will require reducing GRAPE's communication complexity, while multiple service collectives will require integration with a services model.

SDA<sub>M</sub> best satisfied the evaluation criteria, meeting all criteria for nineteen independent variable combinations (see Table 10). The success rate, communication, and utility criteria were satisfied for all independent variable combinations. The runtime requirement was also met with 25–500 robots, and relaxing the requirement to < 10 min enables SDA<sub>M</sub> to satisfy the criteria with 1000 robots and a single service type. However, SDA<sub>M</sub>'s runtimes with five service types and 1000 robots were too long for near real-time domains.

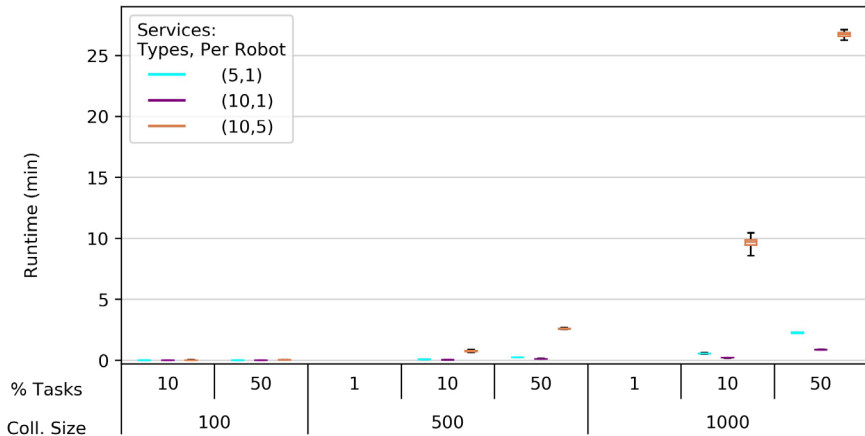
Overall, SDA<sub>M</sub> satisfied the domain criteria reasonably well for single service collectives, and GRAPE produced solutions even more quickly, given sufficient communication infrastructure. However, no algorithm fully satisfied the performance criteria with five service types and 1000 robots.

### 6.3 Heterogeneous Collective Results

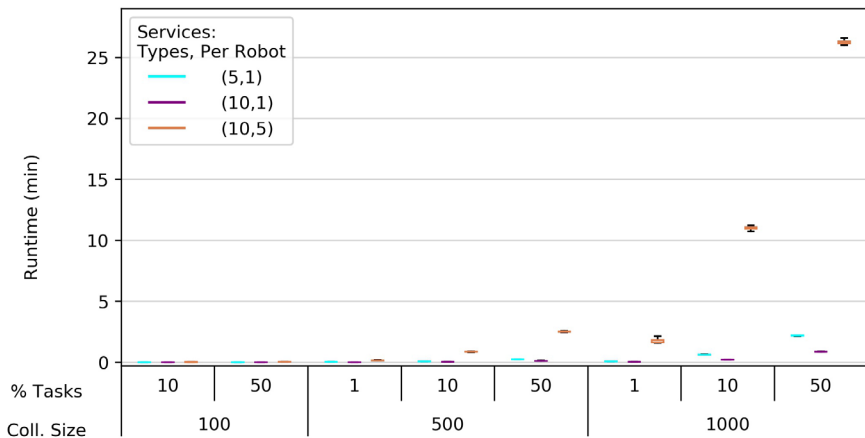
A total of 900 heterogeneous collective trials were run for each of RACH-NA, RACHNA<sub>dt</sub>, and SDA<sub>M</sub>. SDA<sub>SCO</sub> was run for only the 300 trials with multiple robot systems (i.e., 25–50 robots), due to its exponential computational complexity. Recall that GRAPE was not considered for heterogeneous collectives, as it does not incorporate the services model.

#### 6.3.1 RACHNA and RACHNA<sub>dt</sub>

RACHNA produced solutions for only 82.9% of trials. The 150 trials with 1% tasks were unsuccessful, due to RACHNA's known limitation. The other four unsuccessful trials occurred with 25 robots and 10% tasks. Specifically, there was one unsuccessful trial with five services and one service per robot, two trials with ten services and one service per robot, and one trial with ten services and five services per robot. RACHNA<sub>dt</sub> had a 100% success rate.



(a) **RACHNA**'s runtimes with heterogeneous collectives.



(b) **RACHNA<sub>dt</sub>**'s runtimes with heterogeneous collectives.

**Fig. 6** RACHNA's and RACHNA<sub>dt</sub>'s runtimes (min) with heterogeneous collectives. All multiple robot trials (i.e., 25–50 robots) completed in < 1 s. RACHNA did not solve problems with 1% tasks. Note that the y-axis maximum is 25 min

RACHNA and RACHNA<sub>dt</sub>'s runtimes differed by  $\leq 6$  min, as shown in Fig. 6. Both algorithms' runtimes were low for multiple robot systems (< 1 s) and increased substantially for collectives with ten service types and five services per robot ( $\leq 27$  min 5 s). The runtimes were lower with five service types and one service per robot ( $\leq 2$  min 16 s), as well as ten service types and one service per robot ( $\leq 53$  s). Runtimes also increased with percent tasks, consistent with the algorithms'  $O(mn^2 \log n)$  per iteration computational complexity. Additionally, the runtimes increased with services per robot and decreased service types, as the number of services per robot divided by the service types determines the number of robots that each service agent must consider. The longest runtimes occurred with 1000 robots, ten service types, five services per robot, and 50% tasks.

RACHNA's runtime increase with respect to collective size was significant ( $H(n_{25} = 71, n = 75) = 627.11, p < 0.01$ ), with significant post-hoc analyses ( $p < 0.01$ ). RACHNA's runtime also increased significantly with percent tasks ( $p < 0.01$ ), decreased service types ( $p = 0.02$ ), and services per robot ( $p < 0.01$ ). Note that service type analysis compared five service types and one service per robot to ten service types and one service per robot. Similarly, services per robot analysis compared ten service types and one service per robot to ten service types and five services per robot.

RACHNA<sub>dt</sub>'s runtimes increased significantly across collective sizes ( $H(n = 75) = 637.45, p < 0.01$ ) and tasks ( $H(n = 150) = 150.44, p < 0.01$ ), with significant post-hoc analyses ( $p < 0.01$ ). Significant differences between services ( $p < 0.01$ ) and services per robot ( $p = 0.02$ ) were also identified.

RACHNA's and RACHNA<sub>dt</sub>'s communication had similar trends (see Fig. 7), although RACHNA<sub>dt</sub> required an additional  $< 500$  MB. Communication increased substantially from multiple robot systems ( $\leq 78.62$  MB) to collectives ( $\leq 27,601.14$  MB or 27.6 GB), consistent with the  $O(n^3 m |S|)$  per iteration communication complexity. Communication also increased with percent tasks and services per robot. 1000 robot collectives with ten services, five services per robot, and 50% tasks required the most communication.

These trends were significant. RACHNA's communication increased significantly across collective sizes ( $H(n_{25} = 71, n = 75) = 616.84, p < 0.01$ ), with significant post-hoc analyses ( $p < 0.01$ ). RACHNA's communication also grew significantly with percent tasks ( $p < 0.01$ ) and services per robot ( $p < 0.01$ ). No difference was identified between five and ten service types.

RACHNA<sub>dt</sub> produced similar results. Significant increases were identified across collective sizes ( $H(n = 75) = 626.14, p < 0.01$ ) and percent tasks ( $H(n = 150) = 356.24, p < 0.01$ ), with significant post-hoc analyses ( $p < 0.01$ ). Five and ten services per robot also differed significantly ( $p < 0.01$ ). No difference was identified between five and ten service types.

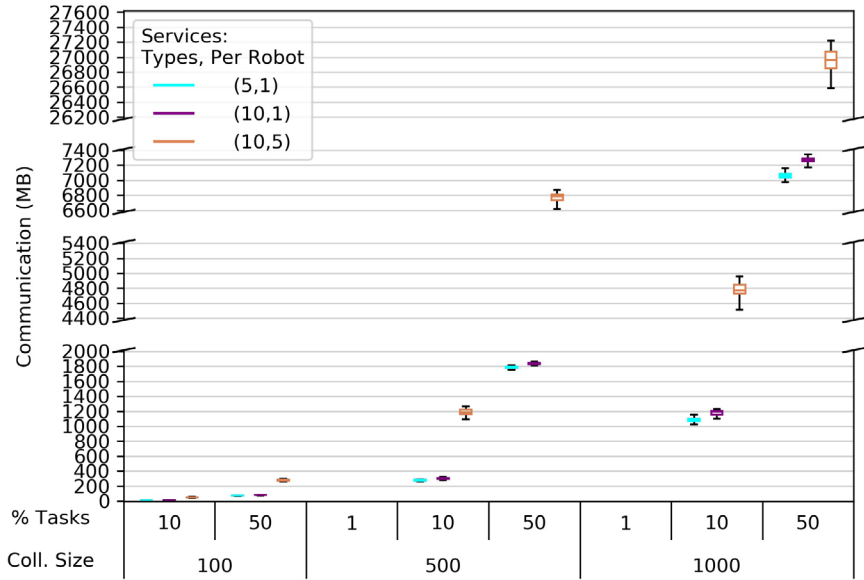
RACHNA produced near-optimal solutions for most problems (Table 11), but high variability caused low worst-case utilities. Utilities were most variable with small collectives and 10% tasks, as with homogeneous collectives.

RACHNA's utilities differed significantly across collective sizes ( $H(n_{25} = 71, n = 75) = 66.49, p < 0.01$ ). Specifically, post-hoc tests found significant differences between 25 and 50–1000 robots ( $p < 0.01$ ), 50 and 500–1000 robots ( $p < 0.01$ ), and 100 and 1000 robots ( $p = 0.03$ ). No significant differences between other collective sizes were identified. 10% and 50% tasks differed significantly ( $p < 0.01$ ), while no significant differences were found between five and ten services, or one and five services per robot. RACHNA<sub>dt</sub> outperformed RACHNA, producing optimal solutions for all trials.

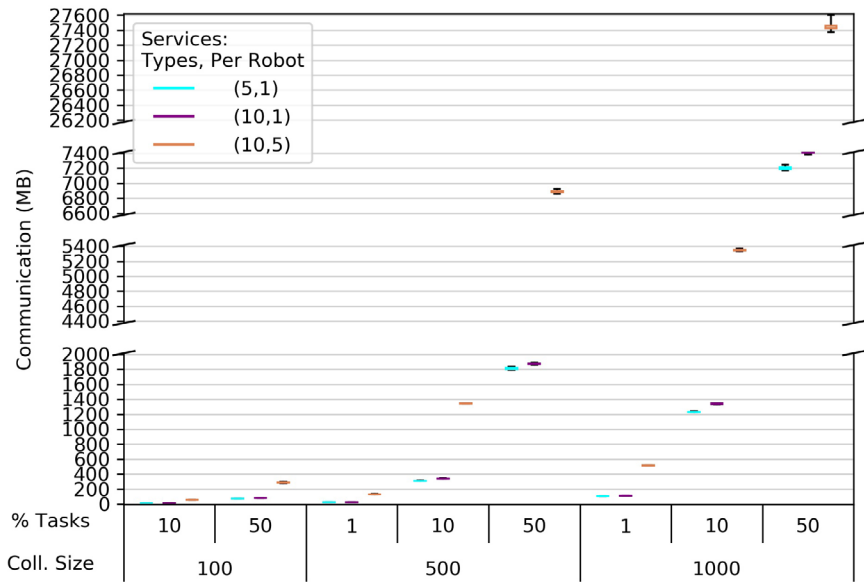
### 6.3.2 Simultaneous Descending Auction

$SDA_M$ 's and  $SDA_{SCO}$ 's results are presented separately, as they differ substantially.  $SDA_M$  had a 100% success rate.

$SDA_M$ 's runtimes (Fig. 8) were low for multiple robot systems ( $< 1$  s), but increased substantially for collectives ( $\leq 29$  min 29 s), consistent with its  $O(mn^4)$  per iteration computational complexity. Additionally, runtimes increased with services per robot, as robots were considered for more roles. 1% and 10% tasks performed similarly, but had higher runtimes than 50% tasks. The decrease in runtime can be attributed to smaller coalitions



(a) RACHNA's communication with heterogeneous collectives.



(b) RACHNA<sub>dt</sub>'s communication with heterogeneous collectives.

**Fig. 7** RACHNA's and RACHNA<sub>dt</sub>'s communication (MB) with heterogeneous collectives. All multiple robot trials (i.e., 25–50 robots) required < 78.62 MB. RACHNA did not solve any problems with 1% tasks. The y-axis maximum is 27,600 MB (i.e., 27.6 GB)

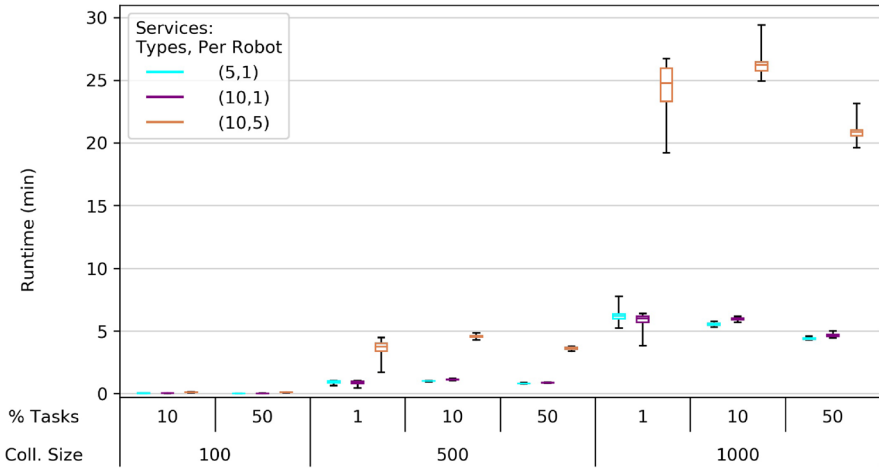
**Table 11** RACHNA's percent utility statistics with heterogeneous collectives

Service Types (Per Robot)	Percent Tasks	Coll Size	RACHNA Median (Min, Max)			
5 (1)	10	25	100.0 (61.29, 100.0)			
		50	93.9 (77.42, 100.0)			
		100	97.56 (85.13, 100.0)			
		500	96.5 (94.59, 98.76)			
	50	1000	1000	95.83 (93.13, 98.31)		
			25	100.0 (99.27, 100.0)		
			50	100.0 (99.51, 100.0)		
			100	99.92 (99.54, 100.0)		
		500	500	99.91 (99.8, 99.98)		
			1000	99.87 (99.81, 99.94)		
			10 (1)	10	25	100.0 (80.43, 100.0)
					50	94.57 (82.54, 100.0)
100	96.85 (87.77, 100.0)					
500	96.6 (93.41, 98.64)					
50	1000	1000		96.26 (93.81, 97.81)		
		25		100.0 (98.39, 100.0)		
		50		100.0 (99.5, 100.0)		
		100		99.92 (99.37, 100.0)		
	500	500		99.89 (99.77, 99.97)		
		1000		99.87 (99.8, 99.94)		
		10 (5)		10	25	100.0 (73.17, 100.0)
					50	97.32 (81.74, 100.0)
100	96.67 (88.99, 100.0)					
500	95.52 (92.47, 98.69)					
50	1000		1000	96.04 (93.37, 97.69)		
			25	100.0 (98.48, 100.0)		
			50	100.0 (99.12, 100.0)		
			100	99.92 (99.28, 100.0)		
	500		500	99.89 (99.71, 99.98)		
			1000	99.89 (99.76, 99.96)		

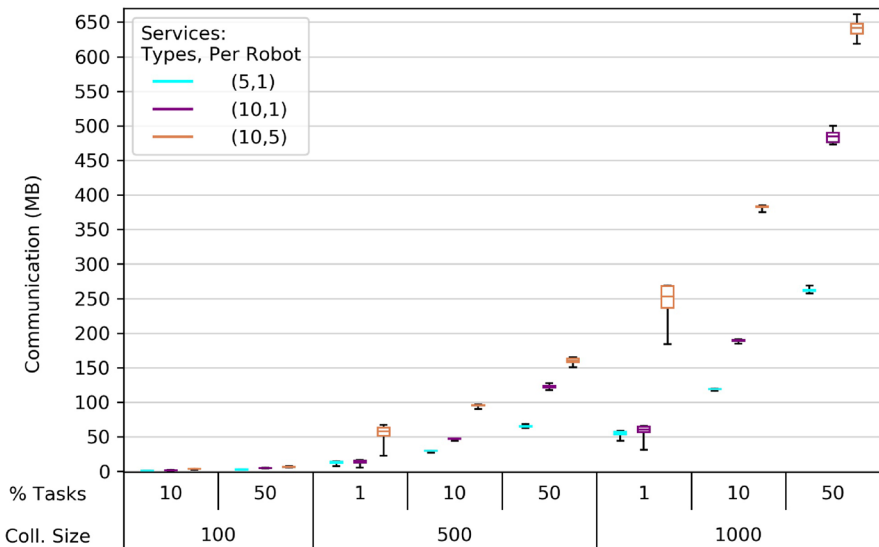
(i.e., higher percent tasks) being cheaper, thus requiring fewer bidding iterations. There was no detectable difference between five and ten service types.

The increases across collective sizes ( $H(n = 75) = 694.31, p < 0.01$ ) and percent tasks ( $H(n = 150) = 24.23, p < 0.01$ ) were significant. Post-hoc tests revealed significant differences between all collective sizes ( $p < 0.01$ ), as well as 1–10% and 50% tasks ( $p < 0.01$ ); however, 1% and 10% tasks did not differ significantly. A significant increase between one and five services per robot was also identified ( $p < 0.01$ ). No difference between five and ten service types was found. The longest runtimes occurred with 1000 robots, ten service types, and five services per robot.

$SDA_M$ 's total required communication (Fig. 9) was < 2 MB with multiple robot systems and increased to as much as 661.39 MB with collectives. The total communication also increased with percent tasks, service types, and service types per robot, consistent



**Fig. 8**  $SDA_M$ 's runtime (min) with heterogeneous collectives. All multiple robot trials (i.e., 25–50 robots) completed in < 1 s. The y-axis maximum is 30 min



**Fig. 9**  $SDA_M$ 's communication (MB) with heterogeneous collectives. All multiple robot trials (i.e., 25–50 robots) required < 2 MB. The y-axis maximum is 650 MB

with  $SDA_M$ 's  $O((m + n)|S|)$  per iteration communication complexity. 1000 robot collectives with ten service types, five services per robot, and 50% tasks required the most communication.

Significant differences were identified across collective sizes ( $H(n = 75) = 693.91, p < 0.01$ ) and percent tasks ( $H(n = 150) = 153.99, p < 0.01$ ), with significant post-hoc analyses ( $p < 0.01$ ). The increases with service types ( $p = 0.04$ ) and services per robot ( $p < 0.01$ ) were also significant.

**Table 12**  $SDA_M$ 's percent utilities with heterogeneous collectives with ten service types, five services per robot, and 1% tasks.  $SA_M$  produced optimal solutions for all other independent variable values

Coll Size	Percent Utility Median (Minimum, Maximum)
25	99.78 (98.29, 100.0)
50	99.94 (99.55, 100.0)
100	99.99 (99.87, 100.0)
500	100.0 (100.0, 100.0)
1000	100.0 (100.0, 100.0)

$SDA_M$  derived optimal solutions with one service per robot, as well as five services per robot and 1% or 10% tasks. The percent utilities with five services per robot and 50% tasks (Table 12) were lower, but still near optimal.

The difference in utility with one and five service types per robot was significant ( $p < 0.01$ ), as was the difference across collective sizes ( $H(n = 75) = 12.39, p = 0.01$ ). Post-hoc tests identified significant differences between collective sizes 25 and 500–1000 ( $p < 0.01$ ), 50 and 500–1000 ( $p = 0.02$  and  $p = 0.03$ , respectively), and 100 and 500–1000 ( $p = 0.03$  and  $p = 0.04$ , respectively). No significant difference was found between other collective sizes, five and ten service types, or across percent tasks.

$SDA_{SCO}$  performed worse than  $SA_M$ , completing only 64.7% (i.e., 194) of the 300 multiple robot system trials. Specifically,  $SDA_{SCO}$  failed to solve any of the 75 trials with 50 robots and 10% tasks, due to exceeding the runtime limit. Additionally,  $SDA_{SCO}$  with ten service types and five services per robot exceeded the time limit for all 25 trials with 25 robots and a 10% tasks, as well as 6 of the 25 trials with 50 robots and a 50% tasks. The remaining multiple robot trials were completed successfully.  $SDA_{SCO}$ 's evaluation did not include the 600 trials with collectives, for an overall 32.3% success rate.

$SDA_{SCO}$ 's runtimes increased with collective size, services per robot, and decreased percent tasks (Table 13). Significant differences were identified between 25 and 50 robot collectives ( $p < 0.01$ ), 10% and 50% tasks ( $p < 0.01$ ), and one and five services per robot ( $p < 0.01$ ). No significant difference existed between five and ten service types. These trends resulted in a maximum runtime (i.e., 1 h 33 min) that is much too long for the near real-time allocation necessary to support high tempo applications, but short enough to support short term pre-mission planning for multiple robot systems.

The communication requirements increased significantly with collective size ( $p < 0.01$ ), as shown in Table 14. As well, the amount of needed communication was significantly

**Table 13**  $SDA_{SCO}$ 's runtimes with heterogeneous multiple robot systems. Collective trials (i.e., 100–1000 robot) were not attempted. No other trials were successful

Service Types	Services Per Robot	Percent Tasks	Coll Size	Runtime (min:s.ms) Median (Minimum, Maximum)
5	1	10	25	1:22:659 (0:08:932, 1:38:643)
			50	0:00:172 (0:00:116, 0:00:442)
			50	0:02:979 (0:01:906, 1:54:099)
10	1	10	25	1:26:369 (0:38:337, 1:44:126)
			50	0:02:185 (0:00:123, 0:01:124)
			50	0:02:887 (0:00:988, 0:14:159)
10	5	50	25	0:05:673 (0:00:679, 16:58:993)
			50	22:50:884 (0:11:952, 93:05:483)



**Table 14**  $SDA_{SCO}$ 's communication with heterogeneous multiple robot systems. Collective trials (i.e., 100–1000 robots) were not attempted. No other trials were successful

Service Types	Services Per Robot	Percent Tasks	Coll Size	Communication (MB) Median (Minimum, Maximum)
5	1	10	25	50.7 (17.05, 63.1)
		50	25	154.63 (101.83, 173.27)
		50	50	639.48 (495.05, 724.42)
10	1	10	25	77.06 (15.98, 99.72)
		50	25	256.22 (200.42, 325.74))
		50	50	1,142.34 (964.76, 1,337.88)
10	5	50	25	384.42 (306.88, 440.69.74)
			50	50

higher based on the percent tasks ( $p < 0.01$ ), services types ( $p < 0.01$ ), and services per robot ( $p = 0.04$ ).

$SDA_{SCO}$  derived optimal solutions for all successful trials with five services and one service per robot, as well as all ten services and one service per robot. The median percent utilities with ten services, five services per robot, and 50% tasks are 99.28 (minimum = 96.0, maximum = 100.0) with 25 robots, and 99.82 (minimum = 99.27, maximum = 100.0) with 50 robots. These utilities are lower than for other compositions, but still near optimal.

## 6.4 Heterogeneous collective discussion

The heterogeneous collective experiments assessed the algorithms' ability to produce near-optimal solutions (i.e., > 95% utility) for very large collectives in near real-time (i.e., < 5 min) and using minimal communication (i.e., < 500 MB), where relaxed thresholds (i.e., < 10 min, < 600 MB, > 90% utility) were considered *reasonably close* to satisfying the criteria. The *hypothesis* that none of GRAPE, RACHNA, RACHNA<sub>dt</sub>,  $SDA_M$ , and  $SDA_{SCO}$  fully satisfied all criteria was supported, even with the relaxed thresholds.

GRAPE was unable to address coalition formation for heterogeneous collectives, as it lacks a services model.  $SDA_{SCO}$  had the next worst performance, meeting all criteria for only five of the thirty-six independent variable combinations (see Table 15). Specifically, all criteria were satisfied for three independent variable combinations with five service types and one service per robot (i.e., 10% tasks and 25 robots, 50% tasks and 25–50 robots), two with ten service types and one service per robot (i.e., 10%-50% tasks and 25 robots), and none with ten service types and one five services per robot. Relaxing the criteria to < 10 min, < 600 MB, and < 90% utility does not enable  $SDA_{SCO}$  to satisfy the criteria for any additional variable combinations.

$SDA_{SCO}$ 's poor performance with multiple robot systems (i.e., 25–50 robots) demonstrated that the algorithm will not scale to collectives.  $SDA_{SCO}$ 's worst case runtimes are far too long for near real-time domains and are expected to increase significantly with collectives. Additionally,  $SDA_{SCO}$ 's  $O(nm^k u_{max}/\epsilon_{dec})$  computational complexity made  $SDA_{SCO}$  very sensitive to variations in coalition size  $k$ , resulting large variations in runtimes. High and variable runtimes, coupled with high communication complexities, prevent  $SDA_{SCO}$  from scaling to collectives.

**Table 15** Heterogeneous results summary:  $SDA_{SCO}$  and RACHNA. Each cell corresponds to a criterion for evaluating viability and an independent variable combination (i.e., an algorithm, services (S), services per robot (R), percent tasks (T), and a collective size). A ✓ means that the algorithm met the criterion for all trials, while a ⚡ means that all trials were reasonably close to meeting the criterion. An ✗ means that the criterion was not met, or trials were not attempted due to known algorithm limitations. A - means that the independent variable combination is invalid

	S R	T	$SDA_{SCO}$					RACHNA				
			Collective Size					Collective Size				
			25	50	100	500	1000	25	50	100	500	1000
Succ. Rate 100%	5/10 1	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✓	✗	✗	✗	✗	✗	✓	✓	✓	✓
		50	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
	10 5	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓
		50	✓	✗	✗	✗	✗	✓	✓	✓	✓	✓
Run- time < 5 min	5/10 1	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✓	✗	✗	✗	✗	✓	✓	✓	✓	✓
		50	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
	10 5	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✗	✗	✗	✗	✗	✓	✓	✓	✓	✗
		50	✗	✗	✗	✗	✗	✓	✓	✓	✓	✗
Comm. < 500 MB	5/10 1	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✓	✗	✗	✗	✗	✓	✓	✓	✓	✗
		50	✓	✓/✗	✗	✗	✗	✓	✓	✓	✗	✗
	10 5	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✗	✗	✗	✗	✗	✓	✓	✓	✗	✗
		50	✓	✗	✗	✗	✗	✓	✓	✓	✗	✗
Utility > 95%	5/10 1	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✓	✗	✗	✗	✗	✗	✗	✗	⚡	⚡
		50	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
	10 5	1	-	-	-	✗	✗	-	-	-	✗	✗
		10	✗	✗	✗	✗	✗	✗	✗	✗	⚡	⚡
		50	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓

RACHNA satisfied all criteria for nine independent variable combinations (i.e., 50% tasks and 25–100 robots, regardless of the collective composition). Relaxing the percent utility criteria to > 90% enables RACHNA to meet all criteria for two additional independent variable combination (i.e., 10% tasks, 500 robots, five or ten service types, and one service per robot). Relaxing the runtime and communication requirements to < 10 min and < 600 MB respectively, does not increase the number of independent variable combinations for which RACHNA satisfies all criteria. RACHNA's primary limitations were the same as with homogeneous collectives, namely high runtimes and communication with 500–1000 robots, as well as an inability to assign coalitions to tasks requiring more robots than their utilities support.

$RACHNA_{dt}$  performed much better than RACHNA, satisfying all criteria for twenty-six independent variable combinations, as shown in Table 16. These independent variable combinations included all those with multiple robot systems, as well as all those with 100 robots. Additionally,  $RACHNA_{dt}$  with five or ten service types and one service per robot satisfied all criteria with 1% tasks and 50–1000 robots, as well as 10% tasks and 500 robots.  $RACHNA_{dt}$  also satisfied all criteria with ten service types, five services per robot, 1% tasks, and 500 robots. Relaxing the criteria does not enable  $RACHNA_{dt}$  to satisfy all criteria for any additional variable combinations.

**Table 16** Heterogeneous results summary: RACHNA<sub>dt</sub> and SDA<sub>M</sub>. Each cell corresponds to a criterion for evaluating viability and an independent variable combination (i.e., an algorithm, services (S), services per robot (R), percent tasks (T), and a collective size). A ✓ means that the algorithm met the criterion for all trials, while a ✓ means that all trials were reasonably close to meeting the criterion. An X means that the criterion was not met, or trials were not attempted due to known algorithm limitations. A - means that the independent variable combination is invalid

	S R	T	RACHNA <sub>dt</sub>					SDA <sub>M</sub>				
			Collective Size					Collective Size				
			25	50	100	500	1000	25	50	100	500	1000
Success Rate 100%	5/10	1	-	-	-	✓	✓	-	-	-	✓	✓
		10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		50	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	10/5	1	-	-	-	✓	✓	-	-	-	✓	✓
		10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		50	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Run-time < 5 min	5/10	1	-	-	-	✓	✓	-	-	-	✓	✓
		10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		50	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	10/5	1	-	-	-	✓	✓	-	-	-	✓	X
		10	✓	✓	✓	✓	X	✓	✓	✓	✓	X
		50	✓	✓	✓	✓	X	✓	✓	✓	✓	X
Comm. < 500 MB	5/10	1	-	-	-	✓	✓	-	-	-	✓	✓
		10	✓	✓	✓	✓	X	✓	✓	✓	✓	✓
		50	✓	✓	✓	X	X	✓	✓	✓	✓	✓/✓
	10/5	1	-	-	-	✓	✓	-	-	-	✓	✓
		10	✓	✓	✓	X	X	✓	✓	✓	✓	✓
		50	✓	✓	✓	X	X	✓	✓	✓	✓	X
Utility > 95%	5/10	1	-	-	-	✓	✓	-	-	-	✓	✓
		10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		50	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	10/5	1	-	-	-	✓	✓	-	-	-	✓	✓
		10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		50	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

RACHNA<sub>dt</sub>'s primary strengths are its high success rates and percent utilities, which, combined with RACHNA<sub>dt</sub>'s low runtimes with one service per robot, mean that the algorithm can be suitable for near real-time domains, given sufficient communication; however, deployed networks will necessitate reducing the communication requirement. Additionally, RACHNA<sub>dt</sub>'s runtimes and communication with five services per robot are not suitable for near real-time domains, especially with 1000 robots.

SDA<sub>M</sub> was the best performing algorithm, meeting all criteria for twenty-nine independent variable combinations (Table 16). Relaxing the runtime criterion to < 10 min and the communication to < 600 MB enabled SDA<sub>M</sub> to satisfy all criteria for four additional trials. The three independent variable combinations for which SDA<sub>M</sub> did not satisfy even the relaxed criteria were ten service types, five services per robot 1%-50% tasks, and 1000 robots.

Overall, SDA<sub>M</sub> met the near real-time domains' coalition formation needs reasonably well for collectives with one service per robot. Additionally, RACHNA<sub>dt</sub> was suitable for collectives with one service per robot, given sufficient communication. However, no algorithm fully satisfied the criteria for very large collectives (i.e., 1000 robots) with five services per robot.

## 7 Discussion

Coalition formation is important for enabling robotic collectives in applications, such as military and disaster response, to perform tasks efficiently. These applications are high-tempo and require operating in highly dynamic domains; thus, high-quality coalition formation solutions must be derived in near real-time (i.e., < 5 min). The applications also typically rely on distributed, low-bandwidth networks, necessitating the use of minimal, distributed communication. A viable collective coalition formation algorithm will be required to satisfy these requirements for very large collectives, across a range of tasks.

This manuscript surveyed existing multiple robot coalition formation algorithms and found that no algorithm was known to scale to very large collectives. Some algorithms were poorly suited to collectives, due to centralization or limited solution spaces, while others' viability was unknown, due to evaluations that considered only < 100 robots or did not consider communication.

Auctions and hedonic games were identified as the most likely to be viable. Some auctions are decentralized and consider heterogeneous systems but had not previously been evaluated with large collectives. Meanwhile, some hedonic games can be distributed and were shown to produce solutions for multiple robot systems in real-time, but did not incorporate a services model.

A simulation-based evaluation assessed auctions' and hedonic games' viability. The auction-based RACHNA, RACHNA<sub>dt</sub>, SDA<sub>M</sub>, and SDA<sub>SCO</sub> algorithms were selected due to their common auction protocols. The hedonic game-based GRAPE algorithm was selected, as it is distributed and most compatible with the services model. The hypothesis that none of the algorithms fully satisfy the coalition formation evaluation criteria (i.e., 100% success rate, < 5 min runtimes, < 500 MB total communication, > 95% utilities) with very

**Table 17** Results summary table. A ✓ means that a viability criterion was satisfied for > 90% of homogeneous (H<sub>mo</sub>) or heterogeneous (H<sub>tr</sub>) independent variable combinations. A ⚡ means that a criterion was satisfied for > 80%, while an ✗ means that a criterion was satisfied for ≤ 80%. Only independent variable combinations with collectives (i.e., > 50 robots) are considered. Independent variable combinations for which an algorithm satisfied a relaxed constraint are counted as 1/2

	SDA <sub>SCO</sub>		RACHNA		RACHNA <sub>dt</sub>		GRAPE		SDA <sub>M</sub>	
	H <sub>mo</sub>	H <sub>tr</sub>	H <sub>mo</sub>	H <sub>tr</sub>	H <sub>mo</sub>	H <sub>tr</sub>	H <sub>mo</sub>	H <sub>tr</sub>	H <sub>mo</sub>	H <sub>tr</sub>
Succ. Rate 100%	✗	✗	✗	✗	✓	✓	✗	✗	✓	✓
Run-time < 5 min	✗	✗	✗	✗	✗	✓	✗	✗	✗	⚡
Comm. < 500 MB	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
Utility > 95%	✗	✗	✗	✗	✓	✓	✗	✗	✓	✓

large collectives was supported, as shown in Table 17. In fact, there were collective compositions for which no algorithm was viable.

Recall that  $SDA_{SCO}$  was unable to satisfy many of the performance criteria, even with smaller multiple robot systems.  $SDA_{SCO}$ 's inability to scale to large multiple robot systems (i.e., 50 robots) demonstrates that the algorithm will not scale to very large collectives (i.e., 1000 robots).

RACHNA's performance was also poor, due to the algorithm's inability to assign large coalitions to lower utility tasks; however,  $RACHNA_{dt}$  addressed this limitation successfully.  $RACHNA_{dt}$ 's strength with both homogeneous and heterogeneous collectives was its high success rates and percent utilities.  $RACHNA_{dt}$  satisfied the runtime criteria for  $< 80\%$  of independent variable combinations with homogeneous collectives, but  $> 90\%$  with heterogeneous collectives. The difference can be attributed to the number of robots that each service agent considered. Homogeneous collectives require all service agents to consider all robots, resulting in additional computation, while heterogeneous collectives only require each service agent to consider a subset of the robots. This observation is consistent with the fact that  $RACHNA_{dt}$  satisfied the runtime criteria for all collectives with one service per robot and not for some collectives with five services per robot. Overall,  $RACHNA_{dt}$  performed well to heterogeneous, single-service collectives, given sufficient communication. However,  $RACHNA_{dt}$ 's communication requirements and its runtimes with other collective compositions were excessive for near real-time domains.

GRAPE performed poorly overall, as it was only able to address homogeneous, single-service collectives. However, GRAPE, when applicable, performed well in terms of success rate, runtime, and percent utility. GRAPE has the potential to be one of the better performing algorithms, if a services model can be integrated without substantially increasing its runtimes. GRAPE's communication requirement will also need to be reduced.

$SDA_M$  was the best performing algorithm, meeting most criteria for  $> 90\%$  of independent variable combinations.  $SDA_M$ 's primary limitation was its long worst case runtimes.  $SDA_M$  had worse runtimes for homogeneous collectives than heterogeneous collectives; however, the underlying cause was the number of services per robot.  $SDA_M$ 's runtimes were longer when there were multiple services per robot, as task agents had to consider each robot for multiple roles, thus requiring more computation. The result was that  $SDA_M$  performed well with homogeneous and heterogeneous collectives of single-service robots, but had runtimes that were too long with multiple-service robots.

These results were generated using a centralized simulator, in which robots' computation was performed iteratively; however, robots in real-world collectives can perform their computation in parallel. All of the algorithms' runtimes are expected to benefit from parallelization. GRAPE is expected to benefit the most, as GRAPE's computation is divided evenly among the robots, while most of the auction algorithms' computation occurs on the task and service agents, which are a small portion of the collective's agents. Parallelization is not expected to substantially impact any algorithms' success rate, communication, or solution quality. Changes to the network topology may have a more substantial impact, as the topology is known to impact GRAPE's performance (Jang et al., 2018), and such changes have not been studied for the auction-based algorithms. Spatial aspects can also impact real-world deployments. The robots' embodiment is addressed through the use of the services model, and robots' positions can be incorporated into the utilities (Jang et al., 2018). However, robots' position in the environment can impact their access to communication (e.g., limited broadcast distance, obstacles blocking network signals), which may

result in lower utilities for all algorithms if robots disconnect from the network during coalition formation.

Overall, no algorithm fully satisfied the performance criteria for very large homogeneous or heterogeneous collectives. Single-service collective coalition formation can be addressed by  $SDA_M$  or, given sufficient communication, GRAPE. Additionally, coalition formation for heterogeneous collectives with single-service robots can be addressed by  $SDA_M$  or, given sufficient communication,  $RACHNA_{dt}$ . However, no algorithm successfully addresses the stated coalition formation performance criteria for homogeneous or heterogeneous collectives with multiple service robots.

Future work must develop faster, low-communication algorithms in order to address these collective compositions.  $SDA_M$  was the best-performing algorithm with other collective compositions, but it is unlikely that modifying  $SDA_M$  will suffice. The underlying bipartite matching algorithm is responsible for much of the computation, limiting the potential runtime gains.  $RACHNA_{dt}$  is similarly limited by the the computation required to determine whether to accept bids, while  $SDA_{SCO}$ 's and  $RACHNA$ 's performance results indicate that they are not well-suited to collectives. These results suggest that a potential path forward is to consider auctions in which computation is distributed more evenly across the collective (i.e., not combinatorial auctions). Modification of GRAPE may also be a viable path forward, if the services model can be incorporated without substantially increasing runtime. This option likely requires reducing GRAPE's communication, although field testing is required to determine how real-world application and network factors (e.g., routing protocols, communication medium, network topology, developments in networking technology, the quantity of communication required for task performance) impact acceptable communication thresholds. It is plausible that the communication reduction can be achieved by relaxing the requirement that only one robot modify the coalition structure during an iteration, thus reducing the number of times communication occurs. A successful near-real time, minimal communication coalition formation algorithm for multiple service collectives will facilitate more diverse and sophisticated collective applications.

## 8 Conclusion

Robotic collectives in the military and disaster response domains will require scalable real-time coalition formation algorithms that produce high quality solutions and use little communication. This manuscript assessed existing algorithms' applicability to homogeneous and heterogeneous collectives with up to 1000 robots and identified distributed hedonic games and auctions as the most promising algorithm types. A simulation-based evaluation of a hedonic game, GRAPE, and auction-based algorithms,  $RACHNA$ , the novel variant  $RACHNA_{dt}$ ,  $SA_M$ , and  $SA_{SCO}$  found that no algorithms consistently met the requirements for collective coalition formation with achievable missions.  $SDA_M$ ,  $RACHNA_{dt}$ , and GRAPE each showed promise with certain collective compositions, but no algorithm fully addressed the needs of collectives with multiple service robots. Future collective coalition formation algorithms will need lower computation and communication in order to address these needs.

**Acknowledgements** Diehl's Ph.D. was supported by an ARCS Foundation Scholar award and the Defense Advanced Research Projects Agency (DARPA). The views, opinions, and findings expressed are those of

the author and are not to be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

**Author contributions** Both authors contributed to the study conception and design. Coding, data collection, and analysis were performed by G.D., supervised by J.A.A.. The first draft of the manuscript was written by G.D., and J.A.A. provided feedback on previous versions of the manuscript. Both authors read and approved the final manuscript.

**Funding** Diehl has received research support from the Defense Advanced Research Projects Agency (DARPA).

**Availability of data and materials** The datasets generated during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** The authors have no conflict of interest to declare outside of funding.

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

## References

- Agarwal, M., Kumar, N., & Vig, L. (2014). Non-additive multi-objective robot coalition formation. *Expert Systems with Applications*, *41*, 3736–3747.
- Aziz, H., Chan, H., Cseh, Á., Li, B., Ramezani, F. & Wang, C. (2021). Multi-robot task allocation—Complexity and approximation. In *International conference on autonomous agents and multiagent systems* (pp. 133–141).
- Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., Viale, M., & Zdravkovic, V. (2008). Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *National Academy of Sciences of the United States of America*, *105*(4), 1232–1237.
- Beni, G. (2005). From swarm intelligence to swarm robotics. In E. Şahin & W. M. Spears (Eds.), *Swarm robotics* (1st ed., pp. 1–9). Springer Berlin Heidelberg.
- Berman, S., Halász, A., Hsieh, M. A., & Kumar, V. (2009a). Optimized stochastic policies for task allocation in swarms of robots. *IEEE Transactions on Robotics*, *25*(4), 927–937.
- Berman, S., Halász, A., Hsieh, M. A., & Kumar, V. (2009b). Optimized stochastic policies for task allocation in swarms of robots. *IEEE Transactions on Robotics*, *25*(4), 927–937.
- Brambilla, M., Ferrante, E., Birattari, M., & Dorigo, M. (2013). Swarm robotics: A review from the swarm engineering perspective. *Swarm Intelligence*, *7*(1), 1–41.
- Castello, E., Yamamoto, T., Nakamura, Y. & Ishiguro, H. (2013). Task allocation for a robotic swarm based on an adaptive response threshold. In *International conference on control, automation and systems* (pp. 259–266).
- Castello, E., Yamamoto, T., Nakamura, Y., & Ishiguro, H. (2014). Foraging optimization in swarm robotic systems based on an adaptive response threshold model. *Advanced Robotics*, *28*(20), 1343–1356.
- Chen, J., & Sun, D. (2011). Resource constrained multirobot task allocation based on leader-follower coalition methodology. *International Journal of Robotics Research*, *30*(12), 1423–1434.
- Clark, S., Usbeck, K., Diller, D., & Schantz, R. E. (2021). CCAST: A framework and practical deployment of heterogeneous unmanned system swarms. *GetMobile: Mobile Computing and Communications*, *24*(4), 17–26.
- Costa, L. F. S., Do Nascimento, T. P., & Goncalves, L. M. G. (2019). Online learning and teaching of emergent behaviors in multi-robot teams. *IEEE Access*, *7*, 158989–159001.
- Couzin, I. D., Krause, J., James, R., Ruxton, G. D., & Franks, N. R. (2002). Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, *218*(1), 1–11.
- Couzin, I. D., Krause, J., Franks, N. R., & Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature*, *433*(3), 513–516.

- Czarnecki, E. & Dutta, A. (2019). Hedonic coalition formation for task allocation with heterogeneous robots. In *IEEE international conference on systems, man and cybernetics* (pp. 1024–1029).
- Czarnecki, E., & Dutta, A. (2021). Scalable hedonic coalition formation for task allocation with heterogeneous robots. *Intelligent Service Robotics*, 14(3), 501–517.
- Defense Advanced Research Projects Agency (2019). OFFensive Swarm-Enabled Tactics. [www.darpa.mil/work-with-us/offensive-swarm-enabled-tactics](http://www.darpa.mil/work-with-us/offensive-swarm-enabled-tactics)
- Diehl, G. & Adams, J. A. (2021a). An ethical framework for message prioritization in disaster response. In *IEEE international symposium on safety, security, and rescue robotics* (pp. 9–14).
- Diehl, G., & Adams, J. A. (2021b). Battery variability management for swarms. In F. Matsuno, S. Azuma, & M. Yamamoto (Eds.), *Distributed autonomous robotic systems* (1st ed., pp. 214–226). Springer.
- Drèze, J. H., & Greenberg, J. (1980). Hedonic coalitions: Optimality and stability. *Econometrica*, 48(4), 987–1003.
- Dutta, A. & Asaithambi, A. (2019). One-to-many bipartite matching based coalition formation for multi-robot task allocation. In *IEEE international conference on robotics and automation* (pp. 2181–2187).
- Dutta, A., Ufimtsev, V., Said, T., Jang, I. & Eggen, R. (2021). Distributed hedonic coalition formation for multi-robot task allocation. In *IEEE international conference on automation science and engineering* (pp. 639–644).
- Elamvazhuthi, K., Biswal, S. & Berman, S. (2018). Mean-field stabilization of robotic swarms to probability distributions with disconnected supports. In *American control conference* (pp. 885–892).
- Gerkey, B. P., & Mataric, M. J. (2002). Sold!: Auction methods for multirobot coordination. *IEEE Transactions on Robotics and Automation*, 18(5), 758–768.
- Gerkey, B. P., & Mataric, M. J. (2004). A formal analysis and taxonomy of task allocation in multi-robot systems. *International Journal of Robotics Research*, 23(9), 939–954.
- Gordon, D. M. (1999). *Ants at work: How an insect society is organized*. Simon and Schuster.
- Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6), 1420–1443.
- Guerrero, J., Oliver, G., & Valero, O. (2017). Multi-robot coalitions formation with deadlines: Complexity analysis and solutions. *PLoS ONE*, 12(1), 1–26.
- Hamann, H. (2018). *Swarm robotics: A formal approach*. Springer.
- Haque, M., Egerstedt, M., & Rahmani, A. (2013). Multilevel coalition formation strategy for suppression of enemy air defenses missions. *Journal of Aerospace Information Systems*, 10(6), 287–296.
- Haque, M., Ren, C., Baker, E., Kirkpatrick, D. & Adams, J. A. (2016). Analysis of swarm communication models. In *Proceedings of the twenty-second European conference on artificial intelligence* (pp. 1716–1717).
- Hartman, C. & Beneš, B. (2006). Autonomous boids. In *Computer animation and virtual worlds* (pp. 199–206).
- Hildmann, H., & Kovacs, E. (2019). Review: Using unmanned aerial vehicles (UAVs) as mobile sensing platforms (MSPs) for disaster response, civil security and public safety. *Drones*, 3(59), 1–26.
- Hsieh, M. A., Halász, A., Berman, S., & Kumar, V. (2008). Biologically inspired redistribution of a swarm of robots among multiple sites. *Swarm Intelligence*, 2(2–4), 121–141.
- Huang, Z. Y. & Robinson, G. E. (1992). Honeybee colony integration: Worker-worker interactions mediate hormonally regulated plasticity in division of labor. In *Proceedings of the national academy of sciences of the United States of America* (pp. 11726–11729).
- Jahir, Y., Atiquzzaman, M., Refai, H., Paranjothi, A., & LoPresti, P. G. (2019). Routing protocols and architecture for Disaster Area Network: A survey. *Ad Hoc Networks*, 82, 1–14.
- Jang, I., Shin, H.-S., & Tsourdos, A. (2018). Anonymous hedonic game for task allocation in a large-scale multiple agent system. *IEEE Transactions on Robotics*, 34(6), 1534–1548.
- Jang, I., Shin, H.-S., Tsourdos, A., Jeong, J., Kim, S., & Suk, J. (2019). An integrated decision-making framework of a heterogeneous aerial robotic swarm for cooperative tasks with minimum requirements. *Journal of Aerospace Engineering*, 233(6), 2101–2118.
- Kanakia, A., Klingner, J., & Correll, N. (2016). A response threshold sigmoid function model for swarm robot collaboration. In N. Chong & Y. Cho (Eds.), *Distributed autonomous robotic systems* (1st ed., pp. 193–206). Springer Japan.
- Kelly, S. M. (2023). Why cell phone service is down in Maui - and when it could be restored. *CNN*, 1–1.
- Klinsompus, P. & Nupairoj, N. (2015). Critical message scheduling for disaster response and recovery phases. In *International conference on information and communication technology convergence* (pp. 65–70).
- Korsah, G. A., Stentz, A., & Dias, M. B. (2013). A comprehensive taxonomy for multi-robot task allocation. *International Journal of Robotics Research*, 32(12), 1495–1512.



- Krieger, M. J., Billeter, J. B., & Keller, L. (2000). Ant-like task allocation and recruitment in cooperative robots. *Nature*, *406*(6799), 992–995.
- Legendre, F., Hossmann, T., Sutton, F. & Plattner, B. (2011). 30 years of ad hoc networking research. In *International conference on wireless technologies for humanitarian relief* (pp. 1–8).
- Liemhetcharat, S., & Veloso, M. (2014). Weighted synergy graphs for effective team formation with heterogeneous ad hoc agents. *Artificial Intelligence*, *208*(1), 41–65.
- Low, K. H., Leow, W. K. & Ang, M. H. (2004). Task allocation via self-organizing swarm coalitions in distributed mobile sensor network. In *AAAI national conference on artificial intelligence* (pp. 28–33).
- Mather, T. W., & Hsieh, M. A. (2011). Macroscopic modeling of stochastic deployment policies with time delays for robot ensembles. *International Journal of Robotics Research*, *30*(5), 590–600.
- Michalak, T. P., Sroka, J., Rahwan, T., Wooldridge, M., McBurney, P. & Jennings, N. R. (2010). A distributed algorithm for anytime coalition structure generation. In *International conference on autonomous agents and multiagent systems* (pp. 1007–1014).
- Mouradian, C., Sahoo, J., Glitho, R. H., Morrow, M. J. & Polakos, P. A. (2017). A coalition formation algorithm for multi-robot task allocation in large-scale natural disasters. In *IEEE international wireless communications and mobile computing conference* (pp. 1909–1914).
- Muralidhar, K. & Madhavi, K. (2018). An investigation into the operational limitations of mobile ad hoc networks. In *IEEE international conference on wireless communications, signal processing and networking* (pp. 1373–1376).
- Oh, G., Kim, Y., Ahn, J., & Choi, H. L. (2017). Market-based task assignment for cooperative timing missions in dynamic environments. *Journal of Intelligent and Robotic Systems: Theory and Applications*, *87*(1), 97–123.
- Pandey, V. K. & De, S. (2017). Communication deployability in disaster management: Taxonomy, recent developments and future challenges. In *IEEE international conference on advanced networks and telecommunications systems* (pp. 1–6).
- Pang, B., Song, Y., Zhang, C., Wang, H., & Yang, R. (2019). Autonomous task allocation in a swarm of foraging robots: An approach based on response threshold sigmoid model. *International Journal of Control, Automation and Systems*, *17*(4), 1031–1040.
- Phelps, S., Cai, K., McBurney, P., Niu, J., Parsons, S., & Sklar, E. (2008). Auctions, evolution, and multi-agent learning. *Adaptive Agents and Multiagent Systems*, *4865*, 188–210.
- Prabhakar, A., Abraham, I., Taylor, A., Schlafly, M., Popovic, K., Diniz, G., Teich, B., Simidchieva, B., Clark, S., & Murphey, T. (2020). Ergodic specifications for flexible swarm control: From user commands to persistent adaptation. In *Robotics: Science and systems* (pp. 1–9).
- Prorok, A., Hsieh, M. A., & Kumar, V. (2017). The impact of diversity on optimal control policies for heterogeneous robot swarms. *IEEE Transactions on Robotics*, *33*(2), 346–358.
- Rahwan, T. & Jennings, N. R. (2008). An improved dynamic programming algorithm for coalition structure generation. In *International conference on autonomous agents and multiagent systems* (pp. 1417–1420).
- Rahwan, T., Ramchurn, S. D., Jennings, N. R., & Giovannucci, A. (2009). An anytime algorithm for optimal coalition structure generation. *Journal of Artificial Intelligence Research*, *34*, 21–567.
- Reina, A., Valentini, G., Fernández-Oto, C., Dorigo, M., & Trianni, V. (2015). A design pattern for decentralised decision making. *PLoS ONE*, *10*, 1–18.
- Rosenfeld, A., Kaminka, G. A., & Kraus, S. (2006). A study of scalability properties in robotic teams. In P. Scerri, R. Vincent, & R. Mailler (Eds.), *Coordination of large-scale multiagent systems* (1st ed., pp. 27–52). Springer.
- Rothkopf, M. H., Pekeč, A., & Harstad, R. M. (1998). Computationally manageable combinatorial auctions. *Management Science*, *44*(8), 1131–1147.
- Sandholm, T., Larson, K., Andersson, M., Shehory, O., & Tohmé, F. (1999). Coalition structure generation with worst case guarantees. *Artificial Intelligence*, *111*(1), 209–238.
- Seeley, T. D. (2010). *Honeybee democracy*. Princeton University Press.
- Sen, S. D. & Adams, J. A. (2013a). A decision network based framework for multiagent coalition formation. In *International conference on autonomous agents and multiagent system* (pp. 55–62).
- Sen, S. D. & Adams, J. A. (2013b). sA-ANT: A hybrid optimization algorithm for multirobot coalition formation. In *IEEE/WIC/ACM international joint conferences on web intelligence and intelligent agent technologies* (pp. 337–344).
- Service, T. C., & Adams, J. A. (2011a). Coalition formation for task allocation: Theory and algorithms. *Autonomous Agents and Multiagent Systems*, *22*(2), 225–248.
- Service, T. C., & Adams, J. A. (2011b). Randomized coalition structure generation. *Artificial Intelligence*, *175*(16–17), 2061–2074.

- Service, T. C., Sen, S. D. & Adams, J. A. (2014). A simultaneous descending auction for task allocation. In *IEEE international conference on systems, man and cybernetics* (pp. 379–384).
- Shah, V. K., Roy, S., Silvestri, S. & Das, S. K. (2019). Towards energy-efficient and robust disaster response networks. In *International conference on distributed computing and networking* (pp. 397–400).
- Shehory, O. & Kraus, S. (1995). Task allocation via coalition formation among autonomous agents. In *International joint conference on artificial intelligence* (pp. 655–661).
- Shehory, O., & Kraus, S. (1998). Methods for task allocation via agent coalition formation. *Artificial Intelligence, 101*, 165–200.
- Sless, L., Hazon, N., Kraus, S. & Wooldridge, M. (2014). Forming coalitions and facilitating relationships for completing tasks in social networks. In *International conference on autonomous agents and multiagent systems* (pp. 261–268).
- Strandburg-Peshkin, A., Twomey, C. R., Bode, N. W. F., Kao, A. B., Ioannou, C. C., Rosenthal, S. B., Torney, C. J., Wu, H. S., Levin, S. A., & Couzin, I. D. (2013). Visual sensory networks and effective information transfer in animal groups. *Current Biology, 23*(17), 709–711.
- Sujit, P. B., George, J. M. & Beard, R. W. (2008). UAV Coalition Formation. In *IEEE American control conference* (pp. 2010–2015).
- Tang, F. (2006). *ASyMTRe: Building coalitions for heterogeneous multi-robot teams*. PhD thesis, University of Tennessee - Knoxville.
- Tang, F. & Parker, L. E. (2005). ASyMTRe: Automated synthesis of multi-robot task solutions through software reconfiguration. In *IEEE international conference on robotics and automation* (pp. 1501–1508).
- Tang, F. & Parker, L. E. (2007). A complete methodology for generating multi-robot task solutions using ASyMTRe-D and market-based task allocation. In *IEEE international conference on robotics and automation* (pp. 3351–3358).
- Theraulaz, G., Bonabeau, E., & Deneubourg, J.-L. (1998). Response threshold reinforcement and division of labour in insect societies. *Proceedings of the Royal Society B: Biological Sciences, 265*(1393), 327–332.
- Van Der Blom, K., & Bäck, T. (2018). A new foraging-based algorithm for online scheduling. In *Genetic and evolutionary computation conference* (pp. 53–60).
- Vig, L., & Adams, J. A. (2006a). Market-based multi-robot coalition formation. In M. Gini & R. Voyles (Eds.), *Distributed autonomous robotic systems* (1st ed., pp. 227–236). Springer.
- Vig, L., & Adams, J. A. (2006b). Multi-robot coalition formation. *IEEE Transactions on Robotics, 22*(4), 637–649.
- Vig, L., & Adams, J. A. (2007). Coalition formation: From software agents to robots. *Journal of Intelligent and Robotic Systems: Theory and Applications, 50*(1), 85–118.
- Xiao, R., & Wang, Y. (2018). Labour division in swarm intelligence for allocation problems: A survey. *International Journal of Bio-Inspired Computation, 12*(2), 71–86.
- Xie, B., Chen, S., Chen, J., & Shen, L. C. (2018). A mutual-selecting market-based mechanism for dynamic coalition formation. *International Journal of Advanced Robotic Systems, 15*(1), 1–10.
- Yeh, C. & Sugawara, T. (2016). Solving coalition structure generation problem with double-layered ant colony optimization. In *International congress on advanced applied informatics* (pp. 65–70).
- Yongming, Y., Xihui, C., Qingjun, L., & Yantao, T. (2010). Swarm robots task allocation based on local communication. *International conference on computer, mechatronics, control and electronic engineering* (pp. 415–418).
- Zahadat, P., & Schmickl, T. (2016). Division of labor in a swarm of autonomous underwater robots by improved partitioning social inhibition. *Adaptive Behavior, 24*(2), 87–101.
- Zahadat, P., Hahshold, S., Thenius, R., Crailsheim, K., & Thomas, S. (2015). From honeybees to robots and back: Division of labour based on partitioning social inhibition. *Bioinspiration and Biomimetics, 10*(6), 066005.
- Zhang, Y., & Parker, L. E. (2012). IQ-ASyMTRe: Forming executable coalitions for tightly coupled multi-robot tasks. *IEEE Transactions on Robotics, 29*(2), 400–416.
- Zhang, Y., & Parker, L. E. (2013). Considering inter-task resource constraints in task allocation. *Autonomous Agents and Multiagent Systems, 26*(3), 389–419.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.