To appear in *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS-15), Hamburg, Germany, September 2015.

Benchmarking Robot Cooperation without Pre-Coordination in the RoboCup Standard Platform League Drop-In Player Competition

Katie Genter[1], Tim Laue[2] and Peter Stone[1]

Abstract— **The Standard Platform League is one of the main competitions of the annual RoboCup world championships. In this competition, teams of five humanoid robots play soccer against each other. In 2014, the league added a new sub-competition which serves as a testbed for cooperation without pre-coordination: the Drop-in Player Competition. Instead of homogeneous robot teams that are each programmed by the same people and hence implicitly pre-coordinated, this competition features ad hoc teams, i. e. teams that consist of robots originating from different RoboCup teams and that are each running different software. In this paper, we provide an overview of this competition, including its motivation and rules. We then present and analyze the results of the 2014 competition, which gathered robots from 23 teams, involved at least 50 human participants, and consisted of fifteen 20-minute games for a total playing time of 300 minutes. We also suggest improvements for future iterations, many of which will be evaluated at RoboCup 2015.**

I. Introduction

As robots become more prevalent in the world, they are increasingly being designed to work in teams to accomplish tasks. Usually, all of the robots on a team are programmed by one organization, and hence are implicitly designed to work together. RoboCup, an annual international robotics competition, features many such teams that are programmed by universities, companies and other organizations to play soccer in various leagues. This paper presents a specific competition held in the Standard Platform League (SPL) at RoboCup 2014, namely the Drop-in Player Competition.

In the Drop-in Player Competition discussed in the paper, each team programmed a robot to coordinate with unknown teammates. The teams were asked not to pre-coordinate, so that during games these agents had to engage in *ad hoc teamwork* in order to reason about their teammates' abilities and intentions in real time and determine how to best assist their team. Each agent's goal was to win a soccer game by as much of a goal difference as possible, while being judged as a 'good teammate' by human officials.

It is often challenging when working with teams of real robots to gather extensive experimental data. The 2014 SPL Drop-in Player Competition gathered robotic agents from 23 teams, involved at least 50 human participants, and consisted of fifteen 20-minute games for a total playing time of 300 minutes. With up to 10 robots participating in each game, this totals to an experiment utilizing up to 50 robot hours. Hence, this competition proved to be the largest ad hoc teamwork experiment on robots that the authors are aware of to date, and is likely one of the largest robotic experiments involving 23 different organizations.

The 2014 SPL Drop-in Player Competition grew from a technical challenge held at RoboCup 2013 [5]. The 2013 technical challenge was optional, and hence only saw six SPL teams participate. Furthermore, the 2013 challenge was announced with little advance notice so many teams did not have time to tailor their strategies to the ad hoc setting. The authors of this paper helped plan, organize, and run the substantially larger SPL Drop-in Player Competition at RoboCup 2014, which was mandatory and announced well in advance. Hence, this paper details the 2014 SPL Drop-in Player Competition, highlights the advancements in the competition as well as in the drop-in player strategies, and suggests improvements for future competitions.

This paper makes two major contributions by (1) presenting the SPL Drop-in Player Competition's setup and rules along with an analysis of its scoring metrics and (2) summarizing and analyzing the participating teams' strategies and comparing their performance in the Drop-in Player Competition with their performance in the main competition. The paper is laid out as follows. Section II describes the SPL as a RoboCup league and introduces the concept of ad hoc teamwork. Details pertinent to the competition are discussed in Section III. The strategies employed by various drop-in players are described in Section IV. Section V presents the results of the 2014 competition, analyzes these results, and suggests some improvements that will be implemented at the 2015 competition. Section VI situates this research in the literature, and Section VII concludes the paper.

II. Background

Two important areas of background knowledge are introduced in this section. The first is the Standard Platform League (SPL) of RoboCup and the second is the multiagent systems research area of ad hoc teamwork.

A. RoboCup Standard Platform League (SPL)

RoboCup is an international robotics competition that had its first competition in 1997. The Standard Platform League (SPL) is different from other RoboCup leagues in that all teams must use the same robotic platform.

Teams in the SPL compete in 5 on 5 soccer games on a 9 meter by 6 meter soccer field, as depicted in Figure 1. Each game consists of two 10-minute halves. Teams must play completely autonomously — no human input is allowed

[1]Katie Genter and Peter Stone are with the Department of Computer Science, University of Texas at Austin, Austin, TX 78712 USA {katie,pstone}@cs.utexas.edu

[2]Tim Laue is with the Department of Computer Science, University of Bremen, 28359 Bremen, Germany tlaue@uni-bremen.de

Fig. 1: NAO robots playing in an SPL game during RoboCup 2014.

during games outside of game state signals sent by an official to communicate to the robots when a goal has been scored, when they have been penalized, etc. The playing environment is color-coded and the robots on each team are allowed to communicate with each other over a wireless network.

B. Ad Hoc Teamwork

Since 1997, RoboCup has served as an excellent domain for testing teamwork, coordination, and cooperation. Most teams have successfully programmed their robots to work well as a team, coordinating which robot should go to the ball, which robot should play defense, and even what formation should be adopted against various opponent types. However, the 2013 drop-in player challenge [5] was one of the first organized efforts to evaluate a player's ability to coordinate with a set of teammates in an ad hoc manner, and the 2014 SPL drop-in player competition greatly improved upon the 2013 challenge in both scale and participation.

Ad hoc teamwork's defining characteristic is its focus on creating agents that can cooperate with unknown teammates without prior coordination. Stone *et al.* imagined 'staging field tests of ad hoc team agents at the annual RoboCup competitions' in their 2010 AAAI challenge paper that introduced ad hoc teamwork [9]. The SPL Drop-in Player Competition at RoboCup 2014 did just this. By organizing the SPL Drop-in Player Competition as a full-fledged competition in which all SPL teams must participate, the authors and the RoboCup organization as a whole have created the potential for a long-standing empirical testbed for ad hoc teamwork research.

III. Competition Description

The Drop-in Player Competition is based on the main RoboCup SPL soccer competition. However, to make it a meaningful competition about teamwork without pre-coordination, several changes are necessary.

A. Altered Rules of the Game

For the most part, the rules of the Drop-in Player Competition games [8] are the same as for main competition games in the SPL. In fact, the only major difference concerns role assignment. In normal SPL games, there is a designated goalkeeper robot on each team. Such a predefined role assignment assigns a particular player to be the goalkeeper instead of forcing the drop-in players to arrange the role assignments for themselves. Hence, in the Drop-in Player Competition, the first robot that enters its own penalty area is considered the goalkeeper for the remainder of the game.

B. Standard Communication

The SPL introduced a wireless standard communication interface[1] in 2014 that is mandatory for all main competition games as well as for all drop-in games. Each message has a predefined format and includes information about the robot's position, walk target, shooting target, observed ball state, and intention. This format enables team-oriented planning by serving as a communication interface.

C. Scoring Scheme

Each player's score in the Drop-in Player Competition consists of two equally weighted components computed over all of the player's games: average goal difference and average judge score. Several important aspects of good team play, such as passing and good position, are not necessarily reflected by a game's final score. As such, each match is also observed by a total of six judges that award positive scores for actions that positively affect team play and negative scores for actions that negatively affect team performance. Judges are given criteria with which to judge the robots. These criteria are detailed in the SPL rulebook [8].

After all drop-in games are complete and the average goal difference and average judge score have been computed for each drop-in player, the two scoring metrics are normalized and added up to determine the overall winner. Specifically, the normalization occurs as follows:

- The player with the highest average goal difference receives 100 points and the player with the lowest average goal difference receives 0 points.
- The player with the highest average human judged score receives 100 points and the player with the lowest average human judged score receives 0 points.
- All other average goal differences and average human judged scores are scaled linearly.
- Each player's judge and goal points are added.

D. Organization of the Competition

In each drop-in game, all 10 robots on the field originate from different teams. To achieve scores that reliably reflect the drop-in capabilities of a single robot, it is best to play as many games as possible with as many different teammates and opponents as possible.

For the RoboCup 2014 Drop-in Competition, we were able to schedule 15 drop-in games. As 25 drop-in players registered, every robot was scheduled to participate in 6 different matches. The pairings were randomly generated by

[1] `https://www.tzi.de/spl/pub/Website/Downloads/SPLStandardMessage.h`

an algorithm that was also used for 3D Soccer Simulation League drop-in games (Algorithm 1 in [5]).

Our schedule was designed with these characteristics:

- Each player did not play with every other player. The algorithm required 27 games in order for each player to play with every other player, which was infeasible.
- During the tournament, each player had at least 18 and up to 22 different teammates.
- Players did not play with each other more than 3 times.
- During the tournament, each player had at least 17 and up to 20 different opponents.

To avoid any pre-coordination, the assignment of robots to teams was announced as late as possible. The time of the announcement varied between 30 minutes and multiple hours, depending on each day's overall schedule. In addition to the late announcement, all participants were explicitly told to refrain from pre-coordinating.

IV. Drop-in Player Strategies

Everyone participating in the 2014 Drop-In Player Competition was asked to submit a short description of the strategy they used in the competition. In total, 17 out of 23 participants submitted a description. The original texts are publicly available [7] in order to allow drop-in players to learn from each other and to provide a better overview of the current status of the competition.

A. Communication and Coordination

As described in Section III-B, all robots within a team are connected by a wireless network and are able to send standardized messages to each other. In theory, the content of these messages should be a valuable source of information when coordinating with teammates. However, in practice, proper communication may not be established because:

- not all robots actually send messages
- not all robots fill in all of the standard message elements
- some robots send incorrect data

In their strategy description, more than half of the participants do not mention these problems or explicitly state that they trust their teammates. However, eight participants mentioned that they do not accept all communicated messages:

- Berlin United, HTWK, and HULKs state that they discard *most* of the information that they receive but they do not discuss how they determine *which* information to discard.
- MiPal did not implement the communication interface.
- Nao Devils and Philosopher send messages but discard all incoming messages.
- B-Human and Northern Bites implemented approaches to determine the reliability of their teammates by checking the plausibility of the transmitted information and the teammates' ability to fulfill their announced roles, respectively. However, Northern Bites did not use this implementation during the competition.

As described in the next section, this limited communication affected the chosen strategies in multiple cases.

B. Typical Player Behaviors

There appears to be one strategy applied by the majority of the drop-in players: *Play the ball if it is close and/or no other robot wants to play the ball. Take a supporting position otherwise.* In many cases, the decision to go to the ball depends on the communicated positions and intentions of teammates. The chosen supporting positions vary from simple strategies like *Stay close to the ball* to more complex computations involving teammate positions. These strategies are, as mentioned by multiple participants, often the same strategies used for their main competition games.

However, some of the participants that accepted limited or no messages from their teammates used a special strategy to avoid conflicts, and thus possible negative scores, with teammates that want to play the ball. They position their robots at fixed positions on the field, e. g. a defensive position inside their own half or somewhere close to the field's center, and wait for the ball to appear nearby. If it does, the robot attempts to kick or dribble towards the opponent goal. Otherwise, it remains at its position and tracks the ball.

One role that was only mentioned in a few descriptions, and rarely seen in games, was the goalkeeper. Some participants actively avoided this role because they believed that the scoring scheme disadvantaged goalkeepers.

C. Suggested Improvements

Drop-in player strategy could be improved in various ways. Firstly, evaluating the trustworthiness of each teammate's communication as well as evaluating the relative skill of each teammate may help determine appropriate behavior. Secondly, finding situations in which it would be reasonable to make or receive a pass could help drop-in players receive better judge scores. Very few drop-in players were witnessed attempting to pass at the 2014 competition. Finally, actively accepting the goalkeeper role may be a good strategy for some drop-in players. This may be especially true when no other teammate has expressed interest in being a goalkeeper and the player has weaker low-level skills than its teammates.

The 2015 competition will be organized to facilitate more useful communication. In particular, it will be checked whether drop-in players are sending complete and reasonable data. Players who are unable to do so will not play until they are able to adequately communicate. Additionally, the communication packet will contain additional fields that allow drop-in players to better convey their abilities.

V. Results and Analysis

The results of the 2014 SPL Drop-in Player Competition were calculated as detailed in Section III-C and are displayed in Table I. This section analyzes these results in detail.

One of the goals of the SPL Drop-in Player Competition is for a team comprised of the top five drop-in players to be able to play comparably to the winner of the main SPL competition. At RoboCup 2014 we held the first of these 'all-star' games where robots from B-Human, HTWK, Nao Devils, TJArk and Berlin United played as a drop-in team against the 2014 SPL main competition champion rUNSWift in a

Team	Country	Judge Avg	Judge Norm	Judge Rank	Goal Diff Avg	Goal Diff Norm	Goal Diff Rank	Drop-in Comp Score	Drop-in Comp Rank	Main Comp Rank	Main vs Drop-in Rank
B-Human	Germany	4.72	100	1	1.33	100	1	200	1	3	+2
HTWK	Germany	1.28	83.04	6	1.00	89.47	2	172.51	2	2	0
Nao Devils	Germany	1.61	84.68	4	0.67	78.95	4-5	163.63	3	5-8	+3.5
TJArk	China	2.17	87.41	3	0.50	73.68	6-7	161.10	4	9-12	+6.5
Berlin United	Germany	-0.58	73.87	12	0.67	78.95	4-5	152.82	5	5-8	+1.5
DAInamite	Germany	0.08	77.15	9	0.50	73.68	6-7	150.84	6	13-20	+10.5
UPennalizers	USA	0.67	80.03	8	0.33	68.42	8-9	148.45	7	9-12	+3.5
Austrian Kangaroos	Austria	-2.90	62.45	19	0.83	84.21	3	146.66	8	9-12	+2.5
rUNSWift	Australia	3.00	91.52	2	-0.17	52.63	13-15	144.15	9	1	-8
Cerberus	Turkey	0.72	80.30	7	0.00	57.89	10-12	138.20	10	13-20	+6.5
Northern Bites	USA	-1.81	67.85	17	0.33	68.42	8-9	136.27	11	13-20	+5.5
NTU RoboPAL	Taiwan	1.61	84.68	4	-0.50	42.11	16-18	126.78	12	5-8	-5.5
Austin Villa	USA	-1.28	70.45	16	-0.17	52.63	13-15	123.08	13	13-20	+3.5
HULKs	Germany	-1.83	67.72	18	-0.17	52.63	13-15	120.35	14	13-20	+2.5
UnBeatables	Brazil	-3.36	60.19	20	0.00	57.89	10-12	118.09	15	-	-
RoboCanes	USA	-1.06	71.55	14	-0.50	42.11	16-18	113.65	16	13-20	+0.5
Philosopher	Estonia	-0.25	75.51	11	-0.67	36.84	19	112.36	17	13-20	-0.5
Edinferno	UK	-0.08	76.33	10	-0.83	31.58	20	107.91	18	13-20	-1.5
MiPal	Australia/ Spain	-0.94	72.09	13	-1.00	26.32	21	98.41	19	-	-
SPQR	Italy	-8.00	37.35	22	0.00	57.89	10-12	95.24	20	9-12	-9.5
MRL	Iran	-1.22	70.73	15	-1.33	15.79	22	86.52	21	5-8	-14.5
UChile	Chile	-4.50	54.58	21	-1.83	0.00	23	54.58	22	4	-18
UTH-CAR	Mexico	-15.6	0.00	23	-0.50	42.11	16-18	42.11	23	-	-

TABLE I: Scores for the 2014 SPL Drop-in Player Competition (listed from best to worst).

full-length game. The result was 4-2 in favor of rUNSWift. rUNSWift's closest game in the main competition was a 5-1 win against HTWK in the championship game. Hence, the fact that the drop-in team was able to take rUNSWift to a 4-2 result is indeed impressive.

Agents designed for the Drop-in Player Competition should be adept at reasoning about their teammates' abilities and intentions and responding in such a way that helps their team the most. The Drop-in Competition scoring metrics, discussed in Section III-C and presented in detail in the SPL rulebook [8], were carefully designed to reward agents for being good teammates and not for just having better low-level skills.

In the following sections we discuss scoring in more detail before comparing results of teams in the main competition with the Drop-in Player Competition and providing some analysis of the winning drop-in players.

A. Judge Scoring

Six human judges watched each Drop-in Player Competition game with the intention of evaluating each player's teamwork abilities. In this section, we discuss why human judges were utilized in this competition before analyzing the judge scoring results, discussing difficulties that we experienced, and presenting the improvements made for the upcoming 2015 competition.

1) Importance: The Drop-in Player Competition is about creating good teammates. Hence, players should be rewarded for good teamwork and not just superior low-level skills. Despite having a standard platform in the SPL, some participants have designed superior walk engines and kick engines that could give them an advantage if only goal difference were considered. Hence, human judges were used to recognize good teamwork that might otherwise be overlooked.

Judge Rank and **Goal Diff Rank** have a Pearson correlation coefficient (R) of 0.3618, meaning they are weakly positively correlated. Lack of a strong correlation implies that judge scoring represents a different quality than goal difference scoring.

2) Analysis: Three columns of Table I refer to judge scores: the **Judge Avg** column, the **Judge Norm** column, and the **Judge Rank** column. The **Judge Avg** column gives the raw average scores given by judges each half. The **Judge Norm** column gives the normalized **Judge Avg**, where the normalization is done as described in Section III-C. Finally, the **Judge Rank** column ranks each drop-in player from best (1) to worst (23) based on its **Judge Norm** score.

When looking at the **Judge Avg** column in Table I, the fact that UTH-CAR had a substantially worse **Judge Avg** than any other drop-in player stands out. UTH-CAR had such a low judge score because their robot was inactive or not on the field for most of its games, and drop-in players that do not appear on the pitch for a half automatically receive a -20 judge score for that half. UTH-CAR's substantially lower **Judge Avg** had a large impact on the results of the competition because it caused 22 of 23 drop-in players to have a **Judge Norm** of greater than 54 and 17 of 23 drop-in players to have a **Judge Norm** of greater than 70. This caused the judge scores to have a weaker influence on the overall Drop-in Player Competition results than desired.

3) Difficulties: It is difficult to design and enforce a scoring scheme for human judging that both fairly assesses teamwork capabilities and is usable for human judges.

During the competition, teams were assigned to provide judges for certain matches. Some teams notified their judge shortly before each game, leaving the judge very little time to become familiar with the judging criteria. Although having the same six judges at each match would provide more consistency, judging duties were distributed across all of the participating teams in order to not place an undue time burden on any particular individuals or teams.

4) Improvements for 2015 Competition: For the 2015 Drop-in Player Competition, we worked to improve upon many of the difficulties experienced at the 2014 competition. We simplified judge scoring by implementing coarser overall criteria and fewer categories for judges to rate. Additionally, an experienced 'head judge' will monitor drop-in player communication and judging for each game and a single page instruction sheet will be available for judges to quickly review. Finally, any players with particularly bad judge scores will be excluded from the judge score linear scaling so as to avoid compressing the other judge scores.

B. Goal Difference Scoring

Each drop-in player's average goal difference was calculated as the average goal difference of the games in which the player was scheduled to compete. In this section we discuss the importance of goal difference in drop-in player scoring, analyze the goal difference scores, discuss some difficulties experienced with goal difference scoring, and present some improvements made for the 2015 competition.

1) Importance: Although judge scoring would ideally determine the best drop-in player, goal difference is also very important because it embodies the main aspect of being a good teammate — helping your team win.

2) Analysis: Three columns of Table I refer to goal difference scores: the **Goal Diff Avg** column, the **Goal Diff Norm** column, and the **Goal Diff Rank** column. The **Goal Diff Avg** column gives the average goal differences across the six games in which the player was scheduled to play. The **Goal Diff Norm** column gives the normalized **Goal Diff Avg**, where the normalization is done as described in Section III-C. Finally, the **Goal Diff Rank** column ranks each drop-in player from best (1) to worst (23) based on its **Goal Diff Norm** score.

Average goal difference across a robot's games is a non-subjective measure of how well the robot's team did. If enough games are run, then average goal difference is a good stand-alone metric for how well a robot plays. However, how well a robot plays is not necessarily a good indication of how good of a *teammate* a robot is.

3) Difficulties: One of the main difficulties that affected average goal differences was that not all drop-in players who registered for the Drop-in Competition showed up. Due to extremely late notice by the missing drop-in players, their spots remained empty which resulted in not all games being played 5 vs. 5. Although this did likely affect the goal difference in these games, we believe the effect was not necessarily significant because their absences were spread across various teammates and opponents.

One of the main difficulties with using a player's average goal difference for determining the best drop-in player is that all players on a team receive the same goal difference from a game despite some players impacting the final game result more than others. In the case of the SPL Drop-in Competition, even players who did not enter a game received credit for the game in terms of goal difference if they were scheduled to play in it. The problem with this is highlighted by the existence of players who missed many games and yet still received better-than-expected goal difference ranks.

4) Improvements for 2015 Competition: For the 2015 Drop-in Player Competition, we have replaced goal difference scoring with a simpler metric of Win, Draw, Loss, or Absent (for registered players who do not participate). This simplification should reduce noise in the scoring process. We are also weighting the game results less than the judge scores when calculating the overall score. Finally, we will play 50% more games than in 2014, but the games will be half games. This will allow more diversity in teammates and opponents.

C. Comparison of Competition Rankings

It can be assumed that teams that perform well in the main SPL RoboCup competition generally have better low-level skills than those who perform poorly. Hence, we can compare each team's Drop-in Player Competition rank (**Drop-in Comp Rank** in Table I) to their main SPL competition rank (**Main Comp Rank** in Table I). These two ranks have a Pearson correlation coefficient (R) of 0.3021, meaning that they are weakly positively correlated. **Main vs Drop-in Rank** in Table I shows how much better or worse each team placed in the Drop-in Player Competition as compared to how it placed in the main competition. Three teams did not compete in the main competition, and hence do not have rankings for that competition.

In general, better teams in the main competition did tend to perform better in the Drop-in Player Competition — only one team that finished tied for 13th in the main competition was in the top 9 teams in the Drop-in Player Competition. Interestingly though, some teams who performed very well in the main competition, namely MRL and UChile, finished in the bottom three teams for the Drop-in Player Competition. This suggests that solid low-level skills and deployment of normal game code will not necessarily yield success in the Drop-in Player Competition.

D. Analysis of the Winning Players

One commonality among four of the top five teams (B-Human, HTWK, Nao Devils, and Berlin United) is their participation in the Drop-In Competition at the RoboCup German Open 2014. This competition was the largest Drop-in Competition test run under realistic conditions prior to RoboCup 2014 — the scenes shown in the supplementary video[2] were recorded at this competition. Furthermore, three

[2] http://www.informatik.uni-bremen.de/agebv2/downloads/videos/genter_laue_stone_iros_15.mp4

of these teams (B-Human, Nao Devils, and Berlin United) participated in the 2013 technical challenge. In this early state of the Drop-In Competition, one can assume that the experiences gained by these teams in these competitions impacted their performance at the RoboCup 2014 competition.

Another noteworthy aspect is the fact that three of the top teams expressed a very limited trust in the communicated messages of others (HTWK and Berlin United) or tried to estimate the reliability of their teammates (B-Human), as described in Section IV-A. This appears to have been one of the insights gained at previous drop-in player competitions: *Avoid cooperation with unreliable teammates*. As mentioned in Section IV-A, currently, it cannot be assumed that all robots comply with the standard communication interface. Indeed, there may be multiple teams with wrong or incomplete implementations that enable them to communicate well with their normal teammates, but poorly with drop-in teammates.

Finally, it should be noted that there was a large variance in the amount of additional code teams implemented for the Drop-in Player Competition. In some cases, developing the teams drop-in player was a team member's main contribution. However, the strategy descriptions [7] also show that some teams just slightly altered their normal game code.

VI. Related Work

Although multiagent teamwork is a well-studied area, most research addresses the problem of coordinating and communicating among teams designed to work together. Ad hoc teamwork, on the other hand, addresses multiagent teamwork in which all of the coordinating agents do not share a common coordination framework. Liemhetcharat and Veloso focused on how to select agents to form ad hoc teams based on each agent's individual characteristics and interactions with its teammates [4]. Barrett *et al.* present empirical evaluations of various types of ad hoc agents when joining coordinated teams of unknown agents [1]. Jones *et al.* present a treasure hunt domain for evaluating ad hoc team performance and present a simple implementation of a team that can search for treasure in such a domain [3].

In the robot soccer domain, Bowling and McCracken [2] propose methods for coordinating an agent that joins an unknown, pre-existing team. In their work, each ad hoc agent is given a playbook that differs from the playbook of its teammates. The teammates assign the ad hoc agent a role, and then react to it as they would any other teammate. The ad hoc agent analyzes which plays work best over hundreds of simulated games, predicts the roles its teammates will adopt in new plays, and assigns itself a complementary role in these new plays. Additionally, the RoboCup Small Size League has held a 'mixed-team' tournament [6] in which two teams are randomly combined to play as one. The winner of their tournament is the team who wins the most games. However, the SPL Drop-in Player Competition is different from this competition in that participation in the SPL competition was much greater and at most one robot from each team joins SPL drop-in games.

Although related, none of this previous research besides the 2013 drop-in challenge [5] has been evaluated on real robots programmed by various organizations from around the world in a truly ad hoc teamwork setting. This paper expands on the 2013 drop-in challenge by substantially increasing both the number of teams participating and the number of drop-in games held.

VII. Conclusion

The SPL Drop-in Player Competition made great strides in becoming a useful testbed for cooperation without precoordination. With the SPL being a standard platform league, and with options existing for teams to just compete in the SPL Drop-in Competition at RoboCup, this testbed is open and approachable for multiagent systems researchers looking to work on ad hoc teamwork in a robotics domain. The authors of this paper, as well as the SPL as a whole, plan to continue this competition for the foreseeable future.

This paper reports on the 2014 SPL Drop-in Player Competition as well as improvements for future competitions. We expect that as teams reflect on this year's results and continue to work on the competition, player strategies will continue to improve. Over time, we expect research towards ad hoc teamwork in the Drop-in Player Competition to flourish alongside the growth of this new domain.

Acknowledgments

Katie Genter and Peter Stone are part of the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (CNS-1330072, CNS-1305287), ONR (21C184-01), AFRL (FA8750-14-1-0070), and AFOSR (FA9550-14-1-0087).

References

[1] S. Barrett, P. Stone, and S. Kraus. Empirical evaluation of ad hoc teamwork in the pursuit domain. In *Proceedings of the Tenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS'11)*, Taipei, Taiwan, May 2011.

[2] M. Bowling and P. McCracken. Coordination and adaptation in impromptu teams. In *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI'05)*, Pittsburgh, PA, USA, July 2005.

[3] E. Jones, B. Browning, M. B. Dias, B. Argall, M. M. Veloso, and A. T. Stentz. Dynamically formed heterogeneous robot teams performing tightly-coordinated tasks. In *Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA'06)*, pages 570–575, Orlando, FL, USA, May 2006.

[4] S. Liemhetcharat and M. Veloso. Modeling mutual capabilities in heterogeneous teams for role assignment. In *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'11)*, San Francisco, CA, USA, September 2011.

[5] P. MacAlpine, K. Genter, S. Barrett, , and P. Stone. The RoboCup 2013 drop-in player challenges: Experiments in ad hoc teamwork. In *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'14)*, Chicago, IL, USA, September 2014.

[6] RoboCup Small Size Robot League. Mixed team tournament, 2013. Online: `http://robocupssl.cpe.ku.ac.th/robocup2013:mixed_team_tournament`.

[7] RoboCup Technical Committee. 2014 drop-in player strategies, 2014. Online: `http://www.tzi.de/spl/bin/view/Website/DropinStrat2014`.

[8] RoboCup Technical Committee. RoboCup standard platform league (nao) rule book, 2014. Online: `http://www.tzi.de/spl/pub/Website/Downloads/Rules2014.pdf`.

[9] P. Stone, G. A. Kaminka, S. Kraus, and J. S. Rosenschein. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI'10)*, Atlanta, GA, USA, July 2010.