The RoboCup Logistics League as a Holistic Multi-Robot Smart Factory Benchmark

Tim Niemueller[1], Alexander Ferrein[2], Sebastian Reuter[3], Sabina Jeschke[3], and Gerhard Lakemeyer[1]

***Abstract*— With autonomous mobile robots receiving increased attention in industrial contexts, the need for benchmarks becomes more and more an urgent matter. The RoboCup Logistics League (RCLL) is one specific industry-inspired scenario focusing on production logistics within a Smart Factory. In this paper, we describe how the RCLL allows to assess the performance of a group of robots within the scenario as a whole, focusing specifically on the coordination and cooperation strategies and the methods and components to achieve them. We report on recent efforts to analyze performance of teams in 2014 to understand the implications of the current grading scheme, and derived criteria and metrics for performance assessment based on Key Performance Indicators (KPI) adapted from classic factory evaluation. We reflect on differences and compatibility towards RoCKIn, a recent major benchmarking European project.**

Fig. 1. Teams Carologistics (robots with additional laptop) and Solidus (pink parts) during the RCLL finals at RoboCup 2015 (Hefei, China).

I. Introduction

Autonomous mobile robots receive increasing attention world-wide in industrial contexts with respect to Smart Factories, for example as as cyber-physical systems of Industry 4.0 [1] efforts. A key question that arises – and which is well-known also from many other robotics domains – is how to assess the performance of such systems and what a benchmark could look like.

We briefly present the RoboCup Logistics League (RCLL) in Section II as an industry-inspired competitive scenario focusing on production logistics in Smart Factories. Its goals are to drive research and education in autonomous mobile multi-robot systems in industrial contexts and to serve as a testbed and benchmark for methods in such environments, in particular regarding multi-robot planning and reasoning, scheduling, coordination, and cooperation. It is a medium complex domain of a comprehensible size which offers a feasible compromise for producing meaningful results and at the same time being (financially) accessible to academic teams. The task of the robots is to autonomously maintain and optimize the material flow in a simplified factory.

We further report on recent efforts to evaluate the league and results of recent years and derive new metrics for performance evaluation based on Key Performance Indicators (KPI) in Section III. We have adapted KPIs used in industry to evaluate (classic) factory performance to provide more fine-grained and tunable criteria to determine the value of robots in the RCLL as well as in more general scenarios. The combination of the assessment of an integrated (multi-robot) system with a domain-specific set of KPIs to evaluate the performance by the autonomous robots is what we call *holistic benchmark*.

In Section IV we briefly mention another recent benchmarking approach from the RoCKIn project and describe some differences and where the approaches are compatible. We describe how our holistic approach allows to evaluate robots in a specific industry-inspired scenario as a whole aiming at long-term robot run-time, rather than designing an artificial scenario with several short-term tasks as done in RoCKIn. We summarize and conclude in Section V.

II. RoboCup Logistics League (RCLL)

RoboCup [2] is an international initiative to foster research in the field of robotics and artificial intelligence. It serves as a common testbed for comparing research results in the robotics field. The industry-oriented RoboCup Logistics League[1] (RCLL) tackles the problem of production logistics in a Smart Factory. Groups of three robots have to plan, execute, and optimize the material flow and deliver products according to dynamic orders. Therefore, the challenge consists of creating and adjusting a production plan and coordinate the group of robots [3]. In 2015, the league has changed considerably by introducing machines that physically modify workpieces during production [4]. We briefly describe the gameplay in 2015, for a more detailed description we refer to [5], in which we also characterize the RCLL as a planning domain. The game in 2014 has been described in [6].

The game is split into two major phases. In the *exploration phase* robots need to roam the environment, discover and

[1]Tim Niemueller and Gerhard Lakemeyer are affiliated with the Knowledge-Based Systems Group, RWTH Aachen University, Aachen, Germany; {niemueller, lakemeyer}@kbsg.rwth-aachen.de

[2]Alexander Ferrein is with the Mobile Autonomous Systems & Cognitive Robotics Institute, FH Aachen University of Appl. Sc., Aachen, Germany ferrein@fh-aachen.de

[3]Sebastian Reuter and Sabina Jeschke are affiliated with the Institute Cluster IMA/ZLW & IfU, RWTH Aachen University, Aachen, Germany {reuter, jeschke}@ima-zlw-ifu.rwth-aachen.de

[1]RCLL website: http://www.robocup-logistics.org

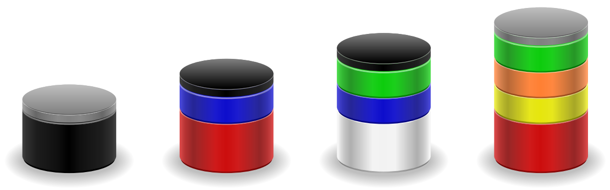

Fig. 2. Examples for products of varying complexity.

identify machines, and report them. Correct reports are rewarded with points, wrong reports reduce points.

In the *production phase*, the goal is to fulfill orders according to a randomized schedule. Orders consist of a specific color-coded product (base element, zero or up to three rings in specific order, and a cap, cf. Fig. 2), and a time window when the product is expected. The robots must then fully autonomously select feasible orders, determine the necessary production steps, and coordinate for an efficient production workflow.

A noteworthy feature is the *referee box* (refbox) [4]. This software component provides agency to the environment by instructing machines and processing their sensor data, communicating with the robots, and posting dynamic orders to the robot teams. The same refbox is also used in a simulation of the RCLL [6], thus the agency of the simulation and of the real-world environment are the same. Additionally, it logs all messages sent over the network (like robot beacon signals with position information), game reports, and all updates to its fact base comprising all information about the game available and relevant for the refbox. This is what allows detailed analyses of RCLL games and what enables benchmarking.

III. RCLL as a Holistic Multi-Robot Benchmark

Besides the goal of the RCLL to drive research and education for mobile multi-robot systems for production logistics in industrial Smart Factory settings, it also aims to be a benchmark for such systems. The RCLL focuses on the evaluation of the efficiency of an overall production process with multi-robot logistics. A major factor enabling this benchmarking is the refbox. As a central communication hub and through its design as a knowledge-based system [4] it allows to record all relevant data. This can then be used to analyze games. This could later also be extended for on-line analysis during games. Benchmarking can mean to play many games with randomized schedules to analyze the system with varying order schedules, or to disable randomization and run with the same order schedule and machine configuration repeatedly, also in simulation.[2]

In this *holistic benchmark* setting, we focus on the value that a group of fully integrated autonomous mobile robots adds to a industrial manufacturing scenario in a Smart Factory as a whole. This in particular evaluates the overall integrated system comprising various functional components, but in particular also the task-level executive. This component is the highest level strategic decision-making component, which determines the (sub-)tasks required and what robots to assign these to. Typical approaches can be roughly divided in three categories: state machine based controllers like SMACH [7] or XABSL [8], reasoning systems from Procedural Reasoning Systems [9] or rule-based agents [10] to more formal approaches like Golog [11], and finally planning systems with varying complexity and modeling requirements, for example based on PDDL [12] and its various extensions.

Another question that arises is how much a single robot contributes to the overall performance of the multi-robot system when added or removed from the team, touching the questions of robustness towards failure and scalability of the integrated system. For example, robots might undergo maintenance and thus be unavailable for a certain time, or the question to be answered is if the productivity can be increased by adding more robots.

Key Performance Indicators

In the following, we will focus particularly on the evaluation of a production logistics scenario, but the basic ideas are applicable to alternative robot scenarios as well.

In [13] we have analyzed 75 GB of data gathered by the refbox during the RoboCup competition in 2014 focusing on the two top performing teams. A key observation made is that the grading scheme implicitly preferred a certain style of production. It meant that minimizing the throughput time of products was valued higher compared to a high machine utilization. This was neither actively intended nor fully understood before this analysis.

This lead to the idea to review Key Performance Indicators (KPI) for production logistics which are metrics used in industrial contexts for evaluating factory performance [13]. We have looked at some well-known KPIs and considered how those could be adapted to evaluate RCLL games. The KPIs considered included the throughput time of products (how long a workpiece needs to pass all processing stations), delivery lateness and reliability to measure deviation from the planned delivery time and the overall adherence to the order schedule, and machines operating in a specific time slice or the overall utilization of individual machines. There is an intrinsic conflict in the context of production logistics known as the scheduling dilemma of logistics [14] that says that some KPIs are in conflict and an appropriate trade-off must be found. For example, there is a conflict between throughput and high machine utilization, the very criteria by which we found the top performing teams in 2014 to differ most.

If accepted into the competition, we expect that the application of KPIs in the context of the RCLL will allow for a much more fine-grained performance evaluation of the overall system of factory environment with agency and the autonomous multi-robot system. It would allow for awarding certain best-in-class performances. It would furthermore enable to define certain criteria (and trade-offs) to model

[2]We used automated scripts to run the simulation system in hundreds of games for the evaluation of agent systems developed during a lab course, cf. https://trac.fawkesrobotics.org/wiki/Projects/LabPRoGrAMR2014

different requirements thus forming a holistic benchmark for a particular production logistics scenario.

While we have focused specifically on multi-robot systems in production logistics, the general idea, to identify specific KPIs for the domain of interest, potentially inspired by a given industrial context, and using such metrics as evaluation criteria, should be applicable to other scenarios.

IV. Discussion

The RoCKIn[3] project is another recent and prominent benchmarking approach with an industrial aspect. For reasons of brevity, we focus on this project for a comparison. There, the goal is to combine system-level and module-level benchmarking results into a single system [15]. The system-level measures the performance of an integrated system for a specific task, while the module-level determines the performance for a specific functional component like manipulation. The approach is to design a competition specifically such that it allows to acquire this information. In contrast, the RCLL starts out with a scenario inspired an industrial production context that aims to automate the production processes to allow for cost-effective production even of small amounts of specific products or of a large number of variants of a product, e.g. according each time to new customer specifications. In the RCLL scenario, it is also relevant to evaluate the effective use of the given production context, e.g., how efficiently the existing machines are used, and how the robots share the overall workload.

While there are key differences between the RCLL and the RoCKIn@Work approach (e.g., single vs. multi-robot, module-level aspects vs. focus on overall integrated system evaluation, scenario design vs. solution design based on industry-inspired scenario), we deem compatible elements in both approaches. RoCKIn's system- and module-level aspects could probably be adapted for the RCLL for specific components, possibly phrased as new KPIs. Likewise, we imagine that applicable KPIs from production scenarios could allow for the evaluation of a robot fleet within the RoCKIn@Work competition. An obstacle in this regard might the difference that RoCKIn@Work focuses on multiple short-term tasks, while the RCLL strives for long-term autonomy by using a single scenario for enduring games or test runs.

V. Conclusion

We have presented the RoboCup Logistics League (RCLL) as a benchmark scenario for production logistics and ideas to improve the performance evaluation based on Key Performance Indicators (KPI) adapted from classic industrial contexts. We have discussed key differences and possible compatibility between RoCKIn as one recent major benchmarking project and the RCLL.

The RCLL focuses on a specific industry-inspired scenario and long-term autonomy and run-time of robots in this environment. This allows to evaluate efficiency, robustness, and scalability of integrated multi-robot systems. While it does provide insights towards the performance of functional components, its focus lies on the coordination of and cooperation within the robot group to improve its efficiency for the Smart Factory logistics task as a whole.

[3]RoCKIn website: `http://rockinrobotchallenge.eu/`

Acknowledgments

T. Niemueller was supported by the German National Science Foundation (DFG) research unit *FOR 1513* on Hybrid Reasoning for Intelligent Systems (`http://www.hybrid-reasoning.org`).

We gratefully acknowledge travel funding provided by Festo Didactic SE to present this work at the open forum on evaluation of results, replication of experiments and benchmarking in robotics research at IROS 2015 in Hamburg, Germany.

References

[1] M. Hermann, T. Pentek, and B. Otto, "Design Principles for Industrie 4.0 Scenarios: A Literature Review," Technische Universität Dortmund, Working Paper 01/2015, 2015.

[2] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa, "RoboCup: The Robot World Cup Initiative," in *Proc. 1st Int. Conf. on Autonomous Agents*, 1997.

[3] T. Niemueller, D. Ewert, S. Reuter, A. Ferrein, S. Jeschke, and G. Lakemeyer, "RoboCup Logistics League Sponsored by Festo: A Competitive Factory Automation Testbed," in *RoboCup Symposium 2013*, 2013.

[4] T. Niemueller, G. Lakemeyer, A. Ferrein, S. Reuter, D. Ewert, S. Jeschke, D. Pensky, and U. Karras, "Proposal for Advancements to the LLSF in 2014 and beyond," in *ICAR – 1st Workshop on Developments in RoboCup Leagues*, 2013.

[5] T. Niemueller, G. Lakemeyer, and A. Ferrein, "The RoboCup Logistics League as a Benchmark for Planning in Robotics," in *WS on Planning and Robotics (PlanRob) at Int. Conf. on Aut. Planning and Scheduling (ICAPS)*, 2015, (to appear).

[6] F. Zwilling, T. Niemueller, and G. Lakemeyer, "Simulation for the RoboCup Logistics League with Real-World Environment Agency and Multi-level Abstraction," in *RoboCup Symposium*, 2014.

[7] J. Bohren and S. Cousins, "The SMACH High-Level Executive," *Robotics Automation Magazine, IEEE*, vol. 17, no. 4, 2010.

[8] M. Loetzsch, M. Risler, and M. Jungel, "XABSL - A Pragmatic Approach to Behavior Engineering," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.

[9] F. Ingrand, R. Chatila, R. Alami, and F. Robert, "PRS: A High Level Supervision and Control Language for Autonomous Mobile Robots," in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 1, 1996.

[10] T. Niemueller, G. Lakemeyer, and A. Ferrein, "Incremental Task-level Reasoning in a Competitive Factory Automation Scenario," in *AAAI Spring Symposium 2013 - Designing Intelligent Robots: Reintegrating AI*, 2013.

[11] H. J. Levesque, R. Reiter, Y. Lespérance, F. Lin, and R. B. Scherl, "Golog: A logic programming language for dynamic domains," *Journal of Logic Programming*, vol. 31, no. 1-3, 1997.

[12] D. McDermott, M. Ghallab, A. Howe, C. Knoblock, A. Ram, M. Veloso, D. Weld, and D. Wilkins, "PDDL – The Planning Domain Definition Language," AIPS-98 Planning Competition Committee, Tech. Rep., 1998.

[13] T. Niemueller, S. Reuter, A. Ferrein, S. Jeschke, and G. Lakemeyer, "Evaluation of the RoboCup Logistics League and Derived Criteria for Future Competitions," in *RoboCup Symposium*, Hefei, China, 2015.

[14] P. Nyhuis and H.-P. Wiendahl, "Logistic production operating curves–basic model of the theory of logistic operating curves," *CIRP Annals-Manufact. Tech.*, vol. 55, no. 1, 2006.

[15] F. Amigoni, A. Bonarini, G. Fontana, M. Matteucci, and V. Schiaffonati, "Benchmarking through competitions," in *European Robotics Forum – Workshop on Robot Competitions: Benchmarking, Technology Transfer, and Education*, 2013.