

# Evaluation of the RoboCup Logistics League and Derived Criteria for Future Competitions

Tim Niemueller<sup>1</sup>, Sebastian Reuter<sup>2</sup>, Alexander Ferrein<sup>3</sup>,  
Sabina Jeschke<sup>2</sup>, and Gerhard Lakemeyer<sup>1</sup>

<sup>1</sup> Knowledge-based Systems Group, RWTH Aachen University, Germany  
{niemueller, gerhard}@kbsg.rwth-aachen.de

<sup>2</sup> Institute Cluster IMA/ZLW & IFU, RWTH Aachen University, Germany  
[firstname.lastname]@ima-zlw-ifu.rwth-aachen.de

<sup>3</sup> MASCOR Institute, Aachen University of Applied Sciences, Germany  
ferrein@fh-aachen.de

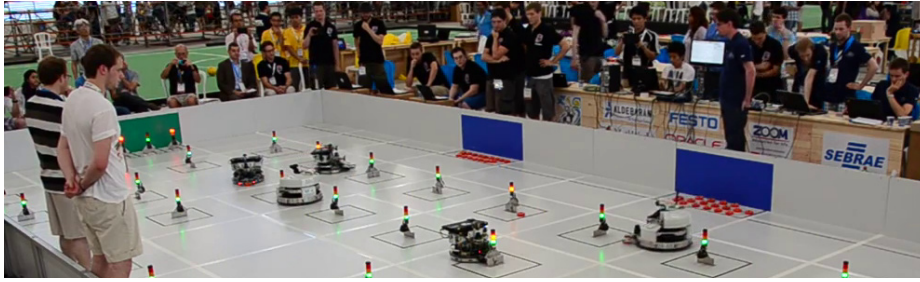
**Abstract.** In the RoboCup Logistics League (RCLL), games are governed by a semi-autonomous referee box. It also records tremendous amounts of data about state changes of the game or communication with the robots. In this paper, we analyze the data of the 2014 competition by means of Key Performance Indicators (KPI). KPIs are used in industrial environments to evaluate the performance of production systems. Applying adapted KPIs to the RCLL provides interesting insights about the strategies of the robot teams. When aiming for more realistic industrial properties with a 24/7 production, where teams perform shifts (without intermediate environment reset), KPIs could be a means to score the game. This could be tried first in a simulation sub-league.

## 1 Introduction

Benchmarking of autonomous mobile robots and industrial scenarios alike are difficult due to many dynamic factors. The scenarios might be too diverse to compare or the environment is not observable (enough). This makes it problematic to evaluate such domains objectively. The *RoboCup Logistics League* (RCLL) is a medium complex domain inspired by actual challenges in industrial applications – in particular that of intra-logistics in a smart factory environment, that is, moving goods in a factory among a number of machines for processing. When developing the league, it was ensured that the domain remained partially observable – enough, so that one could autonomously judge the game.

In an industrial setting, companies strive to improve in terms of *Key Performance Indicators (KPI)*. KPIs are, for example, the time required to move a part through its production process along several machines, or how many products are currently worked on (work in progress) at a time.

Our goal is to *make KPI applicable in the RCLL* in a meaningful way. As a first step, we have analyzed games of the RCLL competition in 2014 focusing on the two top performing teams Carologistics and BBUnits. We provide an evaluation in terms of KPIs mapped to the RCLL game. This is possible, because the



**Fig. 1.** Carologistics (three Robotino 2 with laptops on top) and BavarianBendingUnits (two larger Robotino 3) during the RCLL finals at RoboCup 2014.

referee box, a program that controls and monitors the game, also records relevant data like game state changes and robot communication. The KPIs adapted for the RCLL provide the *performance metrics* by which we can analyze this data. Based on this analysis we give possible explanations on the differences in performance seen from the two teams. The information gained also allows for improving the RCLL as a testbed for industrial applications.

Additionally, on the road to a more realistic industrial setting it is conceivable to aim for a 24/7 production where teams take over shifts without an intermediate environment reset. That would allow for better judging of system robustness and flexibility of the task-level coordination of a team. However, this requires new metrics to score the game, which the adapted KPIs might provide. The RCLL simulation [1] might be a suitable basis to try this in a reasonable way.

In the following Sect. 2, we introduce the RCLL in more detail. In Sect. 3, we give an overview of related work regarding robotic competitions and benchmarks. KPIs and their adaptation to the RCLL is presented in Sect. 4, before applying them for analyzing the RCLL 2014 finale in Sect. 5. We conclude in Sect. 6.

## 2 RoboCup Logistics League

RoboCup [2] is an international initiative to foster research in the field of robotics and artificial intelligence. The basic idea of RoboCup is to set a common testbed for comparing research results in the robotics field. RoboCup is particularly well-known for its various soccer leagues. In the past few years, application-oriented leagues received increasing attention. In 2012, the new industry-oriented RoboCup Logistics League (RCLL, previously LLSF), was founded to tackle the problem of production logistics. Groups of up to three robots have to plan, execute, and optimize the material flow in a smart factory scenario and deliver products according to dynamic orders. Therefore, the challenge consists of creating and adjusting a production schedule and coordinate the group of robots. In the following, we describe the rules of 2014, that we used for our evaluation.

The RCLL competition takes place on a field of  $11.2\text{ m} \times 5.6\text{ m}$  (Fig. 1). Two teams are playing at the same time competing for points, (travel) space and time. Each team has an exclusive input storage (blue areas) and delivery zone (green

area in Fig.1). Machines are represented by RFID readers with signal lights on top indicating the machine state. At the beginning all pucks (representing the products) have the raw material state, are in the input storage, and can be refined (through several stages) to final products using the production machines. These machines are assigned a type randomly at the start of a match which determines what inputs are required and what output will be produced, and how long this conversion will take [1]. Finished products must then be taken to the active gate in the delivery zone. The game is controlled by the referee box (refbox), a software component which instructs and monitors the game [3]. It posts orders dynamically that state the product type (required final puck state), how many items are requested, and a time window when the order must be delivered. Pucks are identified by a unique ID stored on an RFID tag to maintain the puck’s virtual state. After the game is started, no manual interference is allowed, robots receive instructions only from the refbox. Teams receive points for producing complex products, delivering ordered products, and recycling. The RCLL is also very interesting from a planning and scheduling point of view [4].

## 2.1 The Referee Box

Overseeing the game requires tracking of more than 40 pucks and their respective states, watching machine areas of 24 machines to detect pucks that are moved out of bounds, checking for the completion of production steps along the production chain awarding points and keeping a score. This can easily overwhelm a human referee and make the competition hard to understand for the audience. Therefore, we introduced a (semi-) autonomous referee box (refbox) in 2013. It controls and monitors all machines on the field, tracks the score, and provides information for visualization to the audience. The interface for the human referees (e.g., to start or pause the game) is shown in Fig. 2. The refbox communicates with all robots on the field. Some core aspects are listed in the following.

*Control.* The refbox must oversee the game implementing the rules defined in the rule book<sup>4</sup>. For this very purpose it uses the rule-based system CLIPS [5]. This part is responsible for awarding points if the robots accomplished a (partial) task.

*Communication.* It must communicate with the robots on the field to provide information, send orders, and receive reports.

*Representation.* A textual or graphical application is required to visualize the current state of the game and to receive command input from the human referees.

*Interfacing.* The referee box needs to communicate with the programmable logic controller (PLC) which is used to set the light signals and read the RFID sensors.

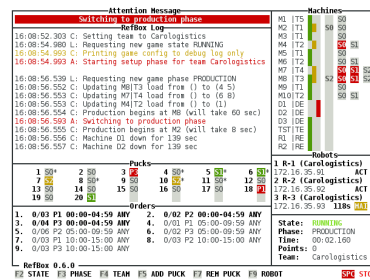


Fig. 2. The Referee Box UI

<sup>4</sup> The current rules can be found at <http://www.robocup-logistics.org/rules>

*Data Recording.* The reffox records each and any message received or sent over the network, all state changes of the internal fact base that is used to control the game, and comprehensive game reports. This is crucial for this work.

### 3 Related Work: Competitions and Benchmarks

Competitions and benchmarking through competitions have become very popular for many research fields from the AI planning and scheduling community (e.g. [6,7]) leading ultimately to the development of PDDL and its extensions over SAT solvers [8] to game-based benchmark for learning algorithms [9] and robotics research. Since its beginnings in the 90's (see [10]), a large number of robotics competitions were launched in all fields of robot applications from autonomous driving (e.g. DARPA Grand Challenges, <http://www.darpa/mil/grandchallenge>) to disaster response (for instance, European Land Robot Trial, <http://www.elrob.org>) to landmine disposal (e.g., Minesweeper, <http://www.landminefree.org>). The motivations for running a competition are manifold. There are aspects to promote or compare research output and approaches. For exchanging ideas and experiences, symposia or user-group meetings are often organized together with a competition to foster the open exchange of solutions and ideas. Additionally, competitions are very motivating and can, in particular, activate students to be part of a competition team.

Among the established robotics competitions, the RoboCup competition [2] is a very successful example. While one of the frequently mentioned motivations of RoboCup is to compare approaches that work well in practice, the comparison of different approaches is nonetheless difficult. A reason is, in part, that robots systems are highly integrated and it is, in general, not possible to exchange software modules or test functionalities in isolation easily. In [11], the authors argue that competition challenges should lead to better algorithms and systems by a continual development process. Anderson et al. [12] critically review the contributions of a number of competitions. Proper benchmarks are not simply given and defined by performing a robotic competition. The organizers of a competition have to define determining factors in order to develop a robotic competition into a benchmark. Many competitions work toward this goal. Under the roof of the RoboCup Federation, in particular, the RoboCup Rescue [13] and RoboCup@Home [14] competitions have to be mentioned. In the RoboCup Rescue competition, for instance, benchmarks for assessing the quality of generated environment maps are established (see e.g. [15]). In RoboCup@Home, the rules change from year to year and an innovative scoring system helps to define a benchmark for fully integrated domestic service robots. Other approaches focus more on certain components such as motion algorithms [16]. The recent RoCKIn project (<http://rockinrobotchallenge.eu/>) aims at setting up a robot competition that increases the scientific and technological knowledge [17,18].

In the next section, we will define the key performance indicators for production systems. These performance indicators can be used in order to judge the

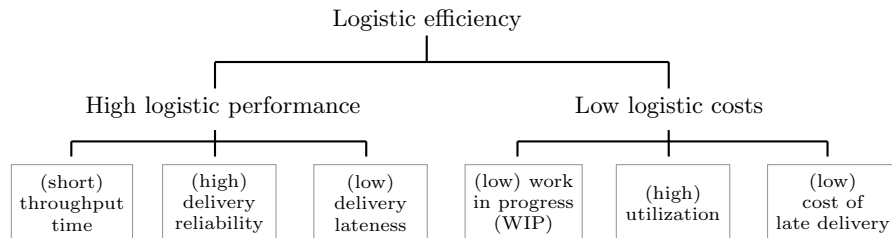
performance of a team. Analyzing the data recorded by the referee box using KPIs, the RCLL could indeed define a logistic benchmark in the future.

## 4 Key Performance Indicators

The traditional goal of production systems (in the sense of systems producing goods, not rule-based production systems) is to maximize production output while minimizing production costs. In the context of increasing market competition, product delivery times and reliability gain importance as buying criteria alongside price and quality of the product [19]. High delivery reliability and short delivery times of products demand for short throughput times of all required intermediate parts and high schedule reliability of all sub-processes within the logistic system [20]. The demand for short throughput time (time span for an order to be created) and high schedule reliability (extent to which planned orders are finished in time) conflicts with the minimization of costs which calls for a high utilization of production resources [21]. Furthermore, the minimization of throughput time and the maximization of output rate contradict each other: As a maximization of output depends on a high level of work in progress (WIP, production orders that are processed in parallel), short throughput times can only be achieved by a low level of WIP [20].

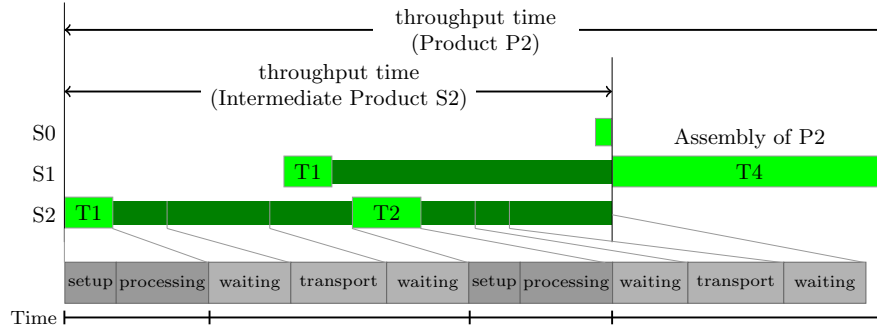
For example, a high utilization of production entities implies a high level of WIP to prevent shortages within the material flow. But it will also slow down the throughput time, because it requires a lot of transport resources. Hence, high machine utilization and short throughput times cannot be achieved together [22].

This conflict among the objectives *logistic performance* and *logistic costs* is called the scheduling dilemma of logistics [22]. Fig. 3 shows Key Performance Indicators as measures for *logistic performance* and *logistic costs* [21]. KPIs are used in industry to make the efficiency of logistic systems assessable.



**Fig. 3.** Key Performance Indicator (KPI) within Production Logistics

The logistic performance can be described by the measures throughput time, delivery reliability and delivery lateness of orders. The *throughput time*  $TTP$  for an operation is defined as start of the order processing ( $T_{\text{operation start}}$ ) till the end of the order processing ( $T_{\text{operation end}}$ ) [21]:  $TTP = T_{\text{operation end}} - T_{\text{operation start}}$ . An exemplary throughput of a product of type P2 is shown in Fig. 4. The production of a product P2 is consists of a manufacture of a intermediate product S1 and



**Fig. 4.** Throughput Time Components

S2. The critical path – the minimal throughput time of a product P2 – is formed by the throughput time of the intermediate product S2 and the final assembly. The manufacturing of an intermediate product S2 consists of two operations on the machines T1 and T2 as well as the time span needed for transportation of the intermediate products (S1 and S2) and the waiting times.

The *delivery lateness*  $DEL$  is a measure for the deviation of the actual ( $T_{\text{actual delivery date}}$ ) and the planned delivery date ( $T_{\text{planned delivery date}}$ ) [21]:  $DEL = T_{\text{actual delivery date}} - T_{\text{planned delivery date}}$ . As the actual delivery of an order can be before and after the specified delivery date, a positive lateness describes an order that was delivered too late and a negative lateness describes an order that was delivered too early. The lateness of an order has a negative impact on the overall delivery reliability of the production system.

The *delivery reliability*  $DERE$  is an indicator to measure if a production systems sticks to scheduled delivery times. It describes the percentage of orders that are delivered within a defined delivery reliability tolerance. The number of in-time deliveries refers to all production orders that are completed within the specified tolerance band of permissible delivery lateness. The number of orders (NO) are all posted orders within the observation period. The *delivery reliability*  $DERE$  can be expressed as [21]:  $DERE = \frac{DEL}{NO} * 100\%$ .

The *logistic costs* influence the effectiveness of a logistics system just as the logistic performance does. As the logistic cost increase, the product price increases and decreases the customers willingness to buy the product. Measure for the logistic costs are work in progress, utilization and cost of late delivery.

The *work in progress*  $WIP$  describes the amount of orders that are started within a production system but are not yet completed. It can be calculated by subtracting the system output from the system input. For discretization, the period of observation can be split into equidistant time slots such as standard hours. Thus, the development of the  $WIP$  can be tracked.

The *utilization*  $U$  describes the ratio of idle and working time of production resources such as production machines or transportation entities. In terms of a production machine the utilization describes the amount of time the machine is processing an item ( $T_{\text{operation}}$ ) in relation to the duration of a reference period ( $T_{\text{duration of reference period}}$ ) [21]:  $U = \frac{\sum T_{\text{operations}}}{T_{\text{duration of reference period}}} * 100\%$ .

The *cost of late delivery COLD* are expenses due to a delay of an order delivery. Cost could be due to increased cost for express shipment or default penalties, or the cost of late delivery can be expressed as a lack of customer trust.

#### 4.1 KPIs applied in the RoboCup Logistics League

The RCLL aims to simulate a realistic, yet simplified, production environment. With the given resources (stationary production machines and mobile robots for transportation) the teams have to maximize the production output with respect to a certain set of products. In this section, we map KPIs to the RCLL.

The *throughput time TTP* in our scenario is defined as the time from the insertion of the first input product (of any accepted type) for a machine until all required inputs have been provided and the processing has been completed. For example, in Fig. 5 in the second line for M2, the Busy-Blocked-Busy cycle is the TTP for a production of 84sec. The *delivery lateness DEL* is directly applicable given that orders have a delivery time window stating a latest time for delivery. The *delivery reliability DERE* can be calculated by dividing the number of delivered products by the total number of products ordered. In the RCLL, *work in progress (WIP)* can be interpreted as machines currently being blocked for an order. This contains machines blocked for the production of intermediate as well as final products, i.e. by the green and orange blocks in Fig. 5. The *utilization U* of a machine is calculated by dividing the actual busy time by the overall game time, i.e. in Fig. 5 all bright green areas. The *cost of late delivery COLD* are expressed in the scoring scheme of the RCLL. A delivery in the requested time window is awarded with 10 points, while a late delivery only scores 1 point, setting COLD to 9 points. Furthermore the RCLL punishes over-production by also reducing the score from 10 to 1 point.

The teams have to balance logistic performance and logistic costs. On the one hand, the teams aim to maximize the logistic performance by short throughput times and low delivery lateness leading to a high output rate and high delivery reliability. On the other hand, the teams have to take care of high WIP which is a prerequisite for high resource utilization, but has a negative effect on throughput time and delivery reliability.

## 5 Analysis of the RCLL 2013 and 2014

For the presented data analysis we have used recordings of the RoboCup competition 2014. The data comprises about 75 GB of rebox data of communication, the state changes of the internal knowledge-based system, text logs, and comprehensive reports of all games played. The data is organized using MongoDB which provides fast and efficient access [23].

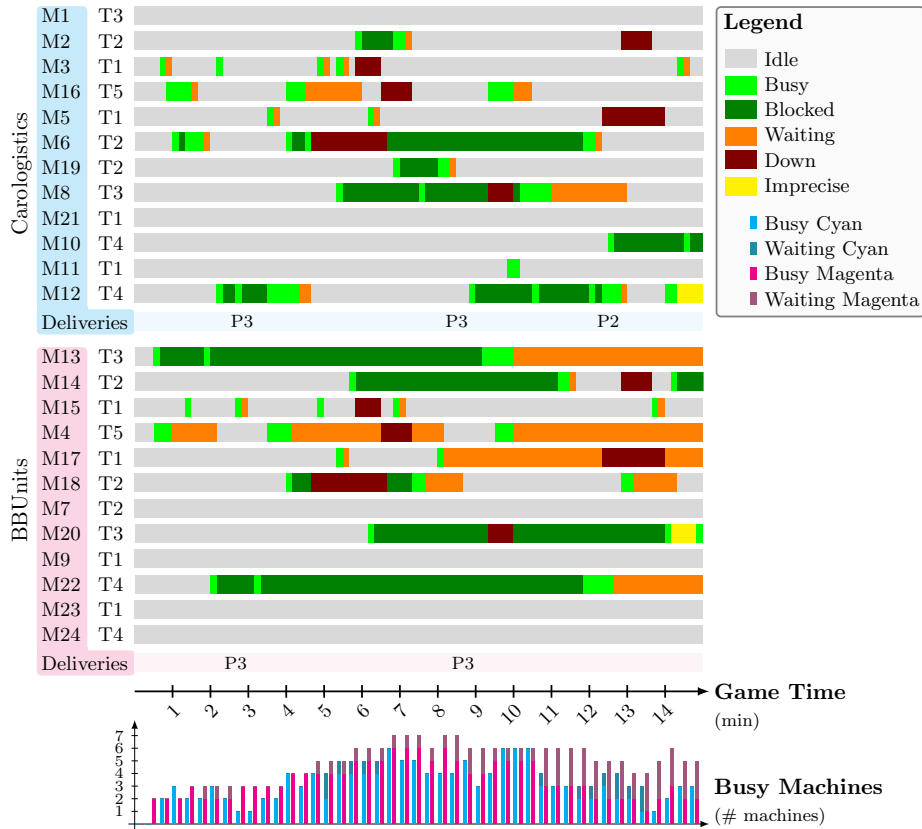
The basic analysis was performed using aggregation and map-reduce features of the database as well as retrieval and analysis scripts written in Python. While we have records for all games of the competitions, for brevity we focus on the two top performers in 2014, the Carologistics and BBUnits teams.

### 5.1 Exemplary application of KPI to the RoboCup 2014 Finale

We will exemplarily apply the KPIs for the RoboCup 2014 final of the RCLL between the Carologistics (cyan) and the BBUnits (magenta) which ended with a score of 165 to 124<sup>5</sup>. We base our analysis on Figs. 5 and 6 for this game.

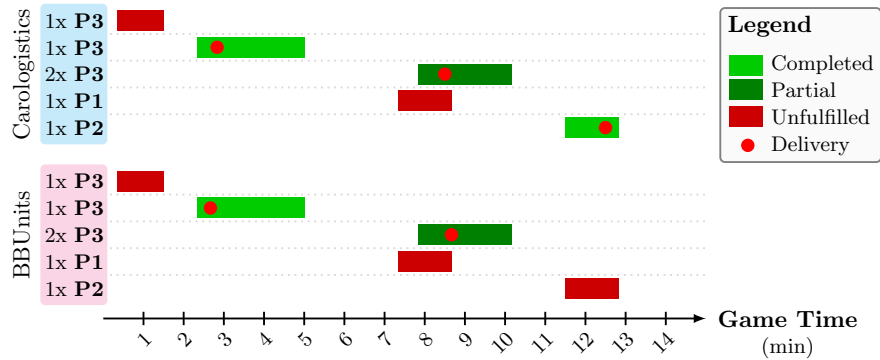
Fig. 5 shows the machines (M1–M24) grouped per team above the time axis. Each row expresses the machine’s state over the course of the game. Gray means it is currently unused (idle). Green means that it is actively processing (busy) or blocked while waiting for the next input to be fed to the machine. After a work order has been completed, the machine is waiting for the product to be picked up (orange). The machine can be down for maintenance for a limited time (dark red). Sometimes the machine is used imprecisely, that is, the product is not placed properly under the RFID device. The row ‘Deliveries’ shows products that are delivered at a specific time. Below the time axis, Fig. 5 shows the busy machines over time. Each entry consists of a cyan and magenta column and

<sup>5</sup> Video of the final is available at [https://youtu.be/\\_iesqH6bNsY](https://youtu.be/_iesqH6bNsY)



**Fig. 5.** Machine states over the course of the final game at RoboCup 2014. The lower graph shows the occupied machines per 20 sec time block.





**Fig. 6.** Adherence to delivery schedule (finals RoboCup 2014). Each row represents an order for the indicated team on the left. The blocks denote their respective delivery windows in the game time represented on the Y-axis. Green boxes mean (partially) completed orders, red unfulfilled ones. Red dots indicate the time of delivery.

represents a 20 seconds period. The height of each column shows the number of machines that are producing in that period (bright team color) or waiting for the product to be retrieved (dark team color).

Fig. 6 shows the orders grouped by teams, which product type was requested and how many were requested. In each row, the colored box denotes the delivery time window in which the product has to be delivered. If the box is green, the order was fulfilled (partial fulfillment means that a smaller number of products was delivered than requested), if it is red no product was delivered in time. The red circles mark the time of delivery. Both teams were able to fulfill the second and third P3 order (partially), but only cyan managed to deliver a P2 product.

The *throughput time TTP* of an order within a machine is denoted by the green (light green and dark green) boxes in Fig. 5. Cyan generally retrieves (partially) finished work orders faster, while magenta often leaves machines blocked for considerable time (dark green areas). The *delivery lateness DEL* can be best seen in Fig. 6. The delivered orders (green boxes) would result in a DEL of zero, while unfulfilled orders would result the maximum DEL of the full game time (in seconds). In some games, orders were delivered after the delivery time window and therefore would get a smaller positive DEL. In the given game, the *delivery reliability DERE* of cyan is 50%, that of magenta is 33%. The *work in progress WIP* machines are shown as busy machines in Fig. 5. As we can see, the WIP was generally equal or higher for magenta, having more machines in use. Looking at the machine states however, most of this time is blocked time in which a machine waits for the next input. If combining with the machines waiting for removal of the finished product, magenta has more machines unusable for new productions on average. The typical *machine utilization U* is currently low in the RCLL due to an emphasis on the logistics aspect that causes long travel times. In the finals, the overall utilization of all machines was about 2.3% by cyan and 1.8% by magenta (thus cyan has utilized the machines more than 25% better). If there had been a late delivery (which did occur in other games), the

*cost of late delivery COLD* would be severe (9 points). What we do see is that some orders were missed completely (resulting in a maximum DEL of losing the full 10 delivery points). In particular, no team managed to fulfill the P1 order. Only cyan managed to complete the work order at all (cf. Fig. 5, row for T3 machine).

**KPI Discussion.** It seems that especially the lower throughput time TTP of cyan contributed to their success. Machines can be used again much faster. For example, only this made it possible to match the delivery time of the P2 order that magenta missed due to very long blocking times. Considering the waiting times makes this even more severe. The cyan team followed the strategy to store products finished before a matching order was received. This meant that the involved machines could be used again much faster (the waiting times of cyan are much shorter). Even with more work in progress machines of magenta, the cyan team used more T1 machines (3 instead of 2). Magenta even left M17 with a finished puck untouched for half of the game. A contributing factor here could be that magenta lost a robot during the game due to a software problem. It seems that the other robots could not recover the state of M17 (instead they later produced at another T1 machine). Concluding it seems that the cyan strategy focused more on low TTP and high throughput, while magenta's strategy was to maximize the overall machine usage and the WIP.

While BBUunits lost a robot in this game, similar statements can be made about a play-offs game between the teams a day earlier that ended 158 to 122 for the Carologistics where both teams had all robots running continuously.<sup>6</sup>

## 5.2 Overall Tournament Evaluation

Analyzing the data of all games at RoboCup2014 (within the Round-Robin Phase, Playoffs and Finals) in terms of machine state graphs (Fig. 5) and adherence to delivery schedules (Fig. 6) as well as using KPIs as statistical queries yields insights for the development of the competition as a whole.

A key insight is that the current dynamic order scheme parameterization is unsuitable for the given resources (robots and machines). Even the best teams at most delivered 3 of 6 ordered products in any game. This seems to be, in particular, because the order time windows are too short. Especially with the modified game in 2015 with vastly more product variants this must be taken into account, since opportunistic production is virtually impossible.

A possible solution would be to considerably increase the time of a game. This would give the robot teams more time to work on the orders and we could gather more data to evaluate the KPIs for a game. It also increases demands for system robustness, a crucial factor for industrial applications. The increased time could be tried first in a simulation league. Work is currently underway to create a common and open simulation for the RCLL based on [1] by the Carologistics and BBUunits team<sup>7</sup>, which could provide the basis for the project.

<sup>6</sup> Video of the play-offs game is available at <https://youtu.be/77V-7LzMBY8>

<sup>7</sup> The project is available at <https://github.com/robocup-logistics>

## 6 Conclusion

In recent years we have developed the RCLL as a domain of medium complexity towards being a testbed for industry-inspired robotic applications. The domain is partially observable by the referee box which allows to record detailed data about the course of the game. This data combined with Key Performance Indicators known from industrial environments allow for analyzing games objectively. We can also use this analysis combined with statistical evaluation to optimize the competition to be more balanced and to improve it as a testbed for industrial robotic applications in smart factory environments.

In an example analysis of the finals in 2014 we have determined some factors based on KPI that may explain the outcome of the game, i.e. that the winning team Carologistics' strategy was focused on short throughput times rather than a high number of machines busy at the same time as the competitor BBUnits did. While we have seen that the order schedule should be tuned to better fit the given resources for more interesting games, teams also need to investigate better scheduling strategies that allow to use the given resources more effectively. KPIs can be one aspect of determining the utility in this regard.

To aim for a more realistic scenario, it is conceivable to develop the RCLL towards a long-time evaluation in the sense of a *24/7 robot competition*. Each team gets assigned a shift in which it has to realize the material flow in the production system without a reset of the environment. Within this scenario a more complex grading scheme is needed as the state of the production system is changing in terms of the amount of work in progress, blocked machines and orders that are currently selected for production. The introduced KPIs are a possible approach to adapt the *grading scheme* to this scenario. It will also require that the teams take different initial states into account and that they provide accurate information to the refbox during a handover to the next team. Especially the development of a *simulation league* can help to facilitate this in a shorter time frame. It would allow teams to adapt more gently. Work in this direction is on-going as described in Sect. 5.2.

More information, the recorded data as well as the evaluation scripts are available at <http://www.fawkesrobotics.org/p/llsf2014-eval>.

## References

1. Zwilling, F., Niemueller, T., Lakemeyer, G.: Simulation for the RoboCup Logistics League with Real-World Environment Agency and Multi-level Abstraction. In: RoboCup Symposium. (2014)
2. Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., Osawa, E.: Robocup: The Robot World Cup Initiative. In: 1st Int. Conference on Autonomous Agents. (1997)
3. Niemueller, T., Lakemeyer, G., Ferrein, A., Reuter, S., Ewert, D., Jeschke, S., Pensky, D., Karras, U.: Proposal for Advancements to the LLSF in 2014 and beyond. In: ICAR – 1st Workshop on Developments in RoboCup Leagues. (2013)
4. Niemueller, T., Lakemeyer, G., Ferrein, A.: 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).

5. Wygant, R.M.: CLIPS: A powerful development and delivery expert system tool. *Computers & Industrial Engineering* **17**(1–4) (1989)
6. Long, D., Kautz, H., Selman, B., Bonet, B., Geffner, H., Koehler, J., Brenner, M., Hoffmann, J., Rittinger, F., Anderson, C.R., et al.: The AIPS-98 planning competition. *AI magazine* **21**(2) (2000)
7. Howe, A.E., Dahlman, E.: A critical assessment of benchmark comparison in planning. *Journal of Artificial Intelligence Research* **17**(1) (2002) 1–33
8. Balint, A., Belov, A., Jarvisalo, M., Sinz, C.: Sat challenge 2012 random sat track: Description of benchmark generation. *SAT CHALLENGE 2012* (2012) 72
9. Karakovskiy, S., Togelius, J.: The mario ai benchmark and competitions. *Computational Intelligence and AI in Games, IEEE Transactions on* **4**(1) (2012) 55–67
10. Balch, T.R., Yanco, H.A.: Ten years of the AAAI mobile robot competition and exhibition. *AI Magazine* **23**(1) (2002)
11. Anderson, M., Jenkins, O., Osentoski, S.: Recasting robotics challenges as experiments [competitions]. *Robotics Automation Magazine, IEEE* **18**(2) (June 2011)
12. Anderson, J., Baltes, J., Cheng, C.T.: Robotics competitions as benchmarks for AI research. *The Knowledge Engineering Review* **26**(01) (2011)
13. Kitano, H., Tadokoro, S.: Robocup rescue: A grand challenge for multiagent and intelligent systems. *AI magazine* **22**(1) (2001)
14. Wisspeintner, T., Van Der Zant, T., Iocchi, L., Schiffer, S.: Robocup@home: Scientific competition and benchmarking for domestic service robots. *Interaction Studies* **10**(3) (2009)
15. Madhavan, R., Lakaemper, R., Kalmár-Nagy, T.: Benchmarking and standardization of intelligent robotic systems. In: *Int. Conf. on Adv. Rob. (ICAR)*. (2009)
16. Calisi, D., Iocchi, L., Nardi, D.: A unified benchmark framework for autonomous mobile robots and vehicles motion algorithms (movema benchmarks). In: *WS on experimental methodology and benchmarking in robotics research at RSS*. (2008)
17. Ahmad, A., Awaad, I., Amigoni, F., Berghofer, J., Bischoff, R., Bonarini, A., Dwiputra, R., Fontana, G., Hegger, F., Hochgeschwender, N., et al.: Specification of general features of scenarios and robots for benchmarking through competitions. *RoCKIn Deliverable D 1* (2013)
18. Amigoni, F., Bonarini, A., Fontana, G., Matteucci, M., Schiaffonati, V.: Benchmarking through competitions. In: *European Robotics Forum–Workshop on Robot Competitions: Benchmarking, Technology Transfer, and Education*. (2013)
19. Nyhuis, P., Wiendahl, H.P.: *Fundamentals of production logistics: theory, tools and applications*. Springer Science & Business Media (2008)
20. Wriggers, F., Busse, T., Nyhuis, P.: Modeling and deriving strategic logistic measures. In: *IEEE Int. Conf. on Ind. Engineering and Engineering Mgmt.* (2007)
21. Lödding, H.: *Handbook of manufacturing control: Fundamentals, description, configuration*. Springer Science & Business Media (2012)
22. Nyhuis, P., Wiendahl, H.P.: Logistic production operating curves–basic model of the theory of logistic operating curves. *CIRP Annals-Manufact. Tech.* **55**(1) (2006)
23. Niemueller, T., Lakemeyer, G., Srinivasa, S.S.: A Generic Robot Database and its Application in Fault Analysis and Performance Evaluation. In: *Int. Conf. on Intelligent Robots and Systems (IROS)*. (2012)