ABSTRACT

In the recent years, the advent of the Industry 4.0, the concepts of Cyber-Physical System and Internet of Things arises, allowing to shift from a classical hierarchical approach to the Manufacturing Planning and Control (MPC) system, to a new class of more decentralised architecture. This paper proposes a decentralised scheduling approach able to improve the performances of a Job-Shop production system, compliant to a semi-heterarchical Industry 4.0 architectures. To this extent, to face with the increasingly complexity of such a scenario, a parametric simulation model able to represent a wide number of Job Shop systems is introduced. Then, through a simulation experimental campaign, the performances of the proposed approach are assessed in function of different control parameter settings. The results showed that the Dispatching Rule Proposed (DRP) led to a significant productivity increase, showing that a semi-heterarchical architecture may be feasible and effective also in a Job-Shop production environment.

1 INTRODUCTION

In recent years, competition in the markets and the complexity of the products require companies to be increasingly flexible. Traditionally, scientific literature differentiates the manufacturing flexibility in both product mix and production volume: the first consists of the production system able to manufacture a large number of different products types; the latter, instead, represents a system able to follow changes in the demand volume while maintaining economic sustainability. Today, these concepts have been further strengthened by the increased willingness of enhanced products customization (Windt and Jeken 2009).

The production world is shifting from a “mass production” scenario, typical of the previous century, towards “mass customisation” one (Fogliatto et al. 2012). Instead of achieving a production focused on a mere cost reduction, it is desirable to create value while meeting customers’ customisation and speed of delivery requirements. The impact of such a strategy is so critical to justifies the born of the “fourth industrial revolution” (also known as “Industry 4.0”) with the introduction of the Cyber-Physical System (CPS) and Internet of Things (IoT) concept. The CPSs are systems in which the “cyber” part, sum of computational and communication capabilities, and the “physical” part are tightly integrated. This brings to a collaborative system of elements linked through the Internet of Things paradigm (Hermann et al. 2016). In particular, the network of CPS, allow to explore different kind of Manufacturing Planning and Control (MPC) approaches, based on the delegation of a defined decision-making quota to the shop floor (Riedl et al. 2014).

The traditional Manufacturing Resource Planning (MRP-II) systems are no longer able to guarantee the ability to adapt to continuously changing conditions due to external or internal reasons (e.g., changes
in demand or plant availability). To this extent, the limit of centralised hierarchical architectures (such as the MRP-II) have been widely investigated in literature (Monostori et al. 2016; Guizzi et al. 2017; Meissner et al. 2017; Moghaddam et al. 2018). Various paradigms based on the shift from a centralised approach (MRP-based) to a decentralised one as a way to cope with more dynamic contexts have been proposed and developed in the literature (Rossit and Tohmé 2018; Dolgui et al. 2019; Ivanov et al. 2018; Sokolov et al. 2018; Dolgui et al. 2018; Ivanov et al. 2018; Rossit et al. 2019). Some of these strategies imply autonomous and independent control concepts, including decision-making methodologies derived from biological examples (Scholz-Reiter 2004; Rekersbrink et al. 2009; Scholz-Reiter et al. 2008).

As shown from these studies, the implementation of a decentralised architecture requires the presence of intelligence distributed among machines and systems on the shop floor, as well as improved communication capability and specific behavioural models. Today, this solution can be implemented thanks to the availability of technologies introduces the Industry 4.0 paradigm. A challenge for decentralised systems is the definition of CPS behavioural models. Typically, in highly decentralised (heterarchical) systems, CPSs may tend to behave in a way such that local and short-term objectives are achieved, rather than aiming to the achievement of system-wide objectives Jeken et al. (2012) (Windt et al. 2008). The problem is that, without global information, decentralised decision-making can converge only to a local optimum rather than a global one, driving to a resource scheduling “reactive” to production disruption. Conversely, a centralised decision-making approach, converging to a global optimum objective, lead to the maximisation of resources utilisation but losing a consistent quota of the required responsiveness, driving to a “proactive” resources scheduling.

To this extent, in literature some studies proposed to overcome these limits with a semi-heterarchical approach: Moghaddam and Deshmukh (2019), Gonzalez et al. (2019), Roa et al. (2019) focused their studies on the physical aspects of the entities of the manufacturing system, while Kuck et al. (2016) concentrated their attention on scheduling and control of production dynamics through data-driven simulation. In contrast, Grassi et al. (2020b) proposed a different approach to the MPC systems in which a defined quota of decision-making autonomy is distributed among different levels, concerning physical, operational and managerial aspects.

In particular, the Grassi et al. (2020b) semi-heterarchical approach is articulated into three levels: Knowledge-based Enterprise Resource Planning (KERP), High-Level Controller (HLC), Low-Level Controller (LLC). The KERP represents an evolution of the classic Enterprise Resource Planning (ERP): it receives orders from the market and makes decisions about their acceptance. It also monitors and controls the Key Performance Indicators (KPIs) of the Production System (PS) by assessing the profitability of orders. The HLC, instead, receives orders from the KERP, monitoring the overall dynamics of the system while balancing the trade-offs between Throughput (TH) and Cycle Time (CT) with a controlling action on the Work-In-Process (WIP) level. Furthermore, the HLC also provides KERP with the KPIs coming from the production system (e.g., CT and TH). Finally, the Low-Level Controller (LLC) is found. It involves the PS and is liable of the short-term scheduling problem. In (Vespoli et al. 2019) and in (Grassi et al. 2020a), the authors faced the problem of the LLC optimisation in the case of Flow-Shop PS, while Guizzi et al. (2019) tackled a similar problem for the Open-Shop systems. In both cases, it is assumed that the production system works with a defined quota of autonomy at the shop floor level and, by comparing different dispatching rules used by the machines, it shows that the use of smart rules for the choice/routing of jobs increases the performance of the production system significantly.

Starting from above-mentioned considerations, this work aims to extend the semi-heterarchical MPC approach to a Job-Shop production system by focusing on proposing a dispatching rule for the LLC level. In order to face the increasing complexity of such a scenario, a parametric simulation model able to represent a wide number of Job Shop systems is introduced. Hence, a new control algorithm for the short-term job-shop scheduling problem, able to increase the performance of the production system while trying to minimise the usually high WIP requirements, is proposed. Finally, an analysis and discussion of the reported result end
the work, showing that a semi-heterarchical architecture may be feasible and effective also in a Job-Shop production environment.

2 Problem Statement

Let us consider a production system organised by departments/production cells that makes custom orders. It is assumed that the production system is composed of 5 departments, each with three alternative machines to carry out the processes assigned to the department. Each department is responsible for a specific operation. Therefore the plant has a total of fifteen machines: department 1 carries out operation 1, department 2 carries out operation 2 and so on up to department 5, which carries out operation 5.

Without losing generality, it is assumed that all the jobs share the same precedence constraints for their production (Figure 1). However, each job has different processing times on each department’s machine, and the operations can be processed within one of the three department’s machine. In addition, each job may start anytime one of its alternative production trees. To be clear, considering the example in Figure 1, in the initial state, the job may start both Operation 1 than Operation 4. Assuming then Operation 1 is completed, the job may start both Operation 2, Operation 3 and Operation 4, and so on. Hence, the production cycle that will be followed by the jobs is not a-priori defined; it will depend on the choices dynamically made at shop floor level.

Classically, production systems are classified into “push” and “pull”. The first ones control the throughput and observe the WIP, while the second ones control the WIP and observe the throughput. (Hopp and Spearman 2011) have shown that pull systems are more efficient than push systems; they achieve the same throughput values with a lower WIP. Here (that is a Job Shop system), a sort of WIP control scheme, like the Flow-Shop CONstant-Work-In-Process (CONWIP), is assumed: the intent is to maintain the WIP controlled within the PS in a similar manner.

As anticipated in the Introduction paragraph, the MPC architecture chosen is the semi-heterarchical one proposed by (Grassi et al. 2020b), where the delivery date of jobs is promised by the KERP on the base of the estimated CT at which the system is working, while the WIP level is appropriately chosen by the HLC level. To this extent, the lower level scheduling (LLC) activities have only to be focused on maintaining the conditions to have TH maximised and CT reduced and stabilized.

In order to test the dispatching rules in a more challenging production environment, we assume that an Operation (Operation 5 in the proposed case-study) represents a bottleneck for most of the jobs. To this extent, the processing time of the first four operations is generated from an exponential distribution with

![Figure 1: Precedence diagram for jobs routing.](image-url)
an average of 10 minutes while the one for the operation 5 comes from an exponential distribution with an average of 20 minutes for the 80% of the jobs, and the remaining 20% from an exponential distribution with an average of 10 minutes.

The choices of the exponential distribution is motivated by the following properties:

- it's a "memory-less" distribution. It means that "it does not remember the past" but behaves at every extraction as if it were the first time. From a probabilistic point of view, this means that \( P(X > x + y | X > y) = P(X > x) \quad \forall x, y; \)
- its variance is significant, and it is equal to the squared average. Then, it is possible to assume that processing time also includes any setup time, machine reconfiguration and transfer from one department to another. Moreover, an exponential distributed processing time, represent the practical worst case possible to be solved (Hopp and Spearman 2011).

3 The proposed approach

As above anticipated, this work aims to extend a semi-heterarchical Manufacturing Planning and Control (MPC) approach to a Job-Shop production system by focusing on proposing a dispatching rule for the lowest machine level (LLC). However, due to the complexity of a Job-Shop Scheduling Problem, we firstly introduce, from a simulation point of view, a parametric model, able to represent a generic Job-Shop system. In particular, its parametric structure allows defining the system dimension in terms of resource number and, most of all, the precedence diagram of a particular job in terms of technological constraints. Then, a newly dispatching rule for the order admittance in a semi-heterarchical Job-Shop environment for the LLC is proposed and discussed.

3.1 Simulation model

A multi-method approach based on Discrete Event Simulation (DES) and Multi-Agent Systems (MAS) was used to develop the simulation model, using the Anylogic 8.5.2 as simulation software. Figure 2 shows the main agent of the proposed model. The Job Ready Queue (JQR) represents a queue that contains the orders released by the KERP. Then, the Production System (PS) accept the order, starting its production, if the number of orders being processed is less than the allowable WIP. Each time a product (Job) leaves the PS, an order is picked from the JQR and inserted into the PS, while the “source” block generates an entity every time one is picked up from JQR. For the sake of clarity, it is assumed that the orders are always available in the JQR.

In Figure 2, two populations of agents are represented: resources and jobs. Resources represent the entities responsible for processing the product (e.g. machines). Jobs are the objects that undergo machining to become, at the end of the technological cycle, finished products. Each job agent contains two additional populations, essential for the work of the Job Shop system: operations and transitions. As expected from a Job-Shop manufacturing system, each job has to respect a precedence diagram for the operations. As clarified in the above paragraph we had chosen to generate jobs with the same precedence diagram, shown in Figure 1, also if the model offers the possibility to associate a different precedence diagram to each job. This diagram - which consists of an adjacency matrix - is represented in the model by means of a population of transitions agents. The information about the resources that can perform the operations, as
Figure 3: Resource State Chart Diagram, alternative resources matrix and adjacency matrix well as the processing times on each resource, are contained in the “operations” population. An example of the adjacency matrix and alternative machines on which operations can be performed is shown in the matrices in Figure 3.

This kind of modelling has high flexibility and makes it possible to represent many job shop problems without having to modify the model and just changing the value matrices showed in Figure 3. The interaction between Job and Resource agents occurs through an evolution of the negotiation mechanism known as Contract Net Protocol (CNP): FIPA Contract Net Interaction Protocol (FIPA-CNIP), which showed promising results for the future application within an Industry 4.0 context. In the proposed model, the negotiation starts with Resource agents because they are usually the least available entity in a production process. When a resource is available (i.e., it is not working) a Call_for_Proposal (CfP) is made to the Jobs that can be worked on that resource.

Therefore the call is made to Jobs that verify the following conditions:

- They are in the “Available” state and, therefore, are not being processed;
- They can be processed by the resource that launched the CfP, according to the precedence diagram and the work already carried out.

The Job agents receive the call, process the reply and return their availability. The job answer consists of a proposal represented by a Proposal agent, which become part of the Proposals population associated with the Job agent. The proposals population contains all the answers elaborated by the Job until the assignment of the Job to the Resource.

The Resource receives the Proposals from the Jobs and keeps them in a different proposals population inside the Resource agent. When all the proposal has been collected, the Resource switches from the “Wait_reply” state to the “Evaluation_Proposal_and_Acceptance” state. In this one, all the proposals are evaluated, assigning a score according to a certain dispatching rule that will be explained later. The proposal that achieves the maximum score is finally chosen and, the Job that sent it, elaborated from the Resource agent.
If the Job accepts the assignment, it switches to the “Execute” state and the Resource switches to the “Execute_task” state. If the Job does not accept the assignment to the Resource - because it has accepted assignments from other resources in the meantime - the negotiation fails, and the Resource returns to the “Available” state. In any case, for both Jobs and Resources, either the negotiation fails or the Job is assigned to the Resource, the proposals associated to the Jobs and Resources are cleared, and the agents are ready for the next negotiation.

If the assignment of the Job to the resource is successful, the Job enters the “Station1” block, located in the resource agent, for machining. This part is modelled in a classical DES (Figure 2). At the end of the machining, the Job and the Resource return to the “Available” state, and the Job records that the operation associated to the resource has been done.

Once all the operations are executed the Job switches to the “Completed” state and exits the production system.

### 3.2 The Dispatching Rule Proposed

As mentioned in the previous paragraph, the resources, during the “Evaluation_and_Acceptance” phase, assign a score to each proposal, choosing the one with the highest score. Here is where the dynamic Dispatching Rule (DRP) takes place, for the score evaluation of the jobs in a Industry 4.0 semi-heterarchical Job-Shop scenario:

\[
Score_{j,m} = X_1 \cdot \frac{PT_{\text{min},j}}{PT_{j,m}} + (1 - X_1) \cdot \frac{C_{\text{System}}}{C_{\text{t}} + PT_{j,m}^{\text{doneOp} + 1}} \tag{1}
\]

where:

- \( j \) is the job that sent the proposal;
- \( m \) is the resource to which the proposal is sent;
- \( X_1 \in [0; 1] \) is a weight value associated with the addend;
- \( PT_{\text{min},j} \) is the minimum processing time of the job \( j \) among all the alternative resources on which it can execute the operation;
- \( PT_{j,m} \) is the processing time of the job \( j \) on the resource \( m \);

Figure 4: Job State Chart Diagram
Table 1: Parameters analyzed in the study scenarios.

<table>
<thead>
<tr>
<th>DispRulePS</th>
<th>FIFO - DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIP</td>
<td>15 - 20 - 25 - 30</td>
</tr>
<tr>
<td>$X_1$</td>
<td>0.6 - 0.8 - 1.0</td>
</tr>
</tbody>
</table>

- $CT_{system}$ is the average Cycle Time of the system (calculated at the time of the proposal evaluation). It is calculated as the average difference value between the Production System exit timestamp and the Production System input timestamp. Therefore this value represents the average value of the sum of the processing time on all resources visited and the waiting time;
- $totOp$ is the number of the planned operations in the technological cycle. In general, it depend on the $j$ job considered. However, in according to the assumptions made in this work, it is equal to 5 for all jobs;
- $CT_j$ is the Cycle Time of the job $j$ until the time of evaluation. It is calculated as the difference between the current simulation timestamp and the Production System input timestamp;
- $doneOp$ is the number of operations already performed on the job $j$.

Analysing it more properly, the first contribute of (1) has the objective of favouring that, the job able to be worked on alternative resources is chosen by the resource where the processing time is shorter. As a matter of fact, in a Job-Shop environment, different machines may complete the same operation with different time (due to a non-necessary setup or because the involved machine is particularly specialised for the given operation). This contribute may show only positive values between $[0; 1]$ and, in the case of no alternative resource for the job, it assumes a value equal to the $X_1$ value. The second contribute of the equation (1), instead, is not limited in its value, and it may assume values higher or lower than 1. The objective, here, is to foster the jobs that show an average Cycle Time for the execution of the operations carried out less than the average one of the system at that step. The control knob of the proposed dispatching rule lies in the $X_1$ value. When it is equal to 1, the rule assumes a behaviour similar to a sort of Short Processing Time (SPT) rule (because the first contribute of the equation is favoured); instead, when it is equal to 0, the rule tends to prefer the job with the lowest crossing time in the system, trying to stabilise the average crossing time of the job within the system. This control knobs will be operated by the future HLC control algorithm, which may change it dynamically, depending on the system performance target.

In order to evaluate the behaviour of the proposed dispatching rules, a wide experimental scenario was analyzed, also comparing the performance of the proposed rule against a sort of “First-In-First-Out (FIFO)” one. To this extent, it should be noted that the FIFO rule is not trivial for a Job Shop system with alternative resources and alternative technological cycles like this one. As a matter of fact, it is appropriate to specify what is meant by FIFO since, in general, it is possible to have overruns between Jobs due to different processing times on resources. In the proposed case study, the FIFO rule has been implemented as follows: when a job agent is generated, it has an incremental ID value; when the negotiation mechanism is triggered during the evaluation phase, the proposal of the Job with the smallest ID will be chosen.

4 Results and Discussion

A campaign of experiments was developed to assess the effectiveness of the proposed approach and, to understand the effect of the following parameters on system performance: Dispatching Rule Production System (DispRulePS), WIP and $X_1$. The values used for the parameters are those indicated in Table 1; the scenarios analysed are related to the experimental plan shown in Table 2.

Each experiment ran for two years (three shifts of 8 hours a day, seven days a week) and was replicated ten times. The Figure 5 shows a comparison between performances obtained with FIFO and DRP rules within different scenarios analyzed. It is clear that, compared to FIFO, DRP always achieves better results in terms of Throughput and Cycle Time - MEAN. In particular, focusing the attention on the $X_1$ parameter,
for the considered scenario, the best PS performance, in term of Throughput and CT - MEAN, were found at a 0.8 value.

In addition, as WIP increases, the DRP performance always shows better performance of the PS system, in terms of Throughput and Cycle Time - MEAN, unlike FIFO. However, it should be noted that the FIFO rule does not show to have any benefits from a WIP value greater than 20. This effect is because, with the use of a FIFO rule, the resources liable of the bottleneck operation results often saturated and the WIP within the production system is concentrated upstream them. A proof of this behaviour is understandable by looking at Figure 6, in which the utilisation of the resources is showed comparing the different value of WIP and Dispatching Rule.

In particular, starting from the 20 WIP value, while using the FIFO rule, the bottleneck resources shows an unitary utilisation. This situation generates starvation for the other resources that, indeed, show an utilisation lower than 56%. A different behaviour may be seen when the DRP takes the scene, in which as the WIP increases, the utilisation of other resources increases too, regardless of the presence of bottleneck resources.

The reason may be found in the (1): the DRP fosters jobs that have a Cycle Time for each operation lower than the average system value and promotes situations in which jobs are assigned to the resource whose processing time is shorter. It means that it favours the better association of jobs and bottleneck resources and that, although the utilisation of bottleneck resources remains unchanged, the DRP tends to associate the jobs to the machine with the lowest processing time, unlike FIFO which associates the jobs to the machine based on the entering timestamp in the production system.

From the other side, the behaviour of the DRP shows a contraindication on the standard deviation of the Cycle Time. From the Figure 5 it is possible to note that as the WIP increases, the Standard Deviation Value increases more than proportionally. The problem is that with this rule, the “slower” jobs may wait a consistent amount of time as a WIP in the Production System, waiting to be processed on the required resource. This effect seems to be admissible for little value of WIP (i.e., up to 20 for the simulated scenario) and to be further investigated for highest value of WIP. However, in order to mitigate this effect, it should be noted that, when the WIP increases, the highest the $X_1$ value, the best the performance is in terms of Cycle Time - DEVIATION are achieved from the system.
5 Conclusions

In this new dynamic mass customisation context, the ability to allocate effectively the available productive resource becomes essential. To this extent, seizing the technological innovations that the Industry 4.0 made available, this work proposed a first decentralised scheduling approach for a job-shop production system. Due to the complexity of such a scenario, a parametric simulation model that allow representing many Job-Shop production system was introduced. As shown, the advantages of this modelling approach lies in its flexibility. In fact, by modifying only some input data, it allow to analyse scenarios related to Job Shop systems without any structural changes to the model.

From the analysis of results, a significant productivity increase in all the considered scenario has been observed with the use of the proposed dispatching rule. In particular, the registered results show the potential of the DRP when using small level of WIP within the production system, compliant with the lean manufacturing approach to the production. Furthermore, in terms of Cycle Time, the proposed approach facilitates significantly the stabilisation of the system, increasing its predictability.

In conclusion, this work may be considered as a first step towards the development of a different approach to the scheduling problem for the job-shop production system, ready to be implemented in the future semi-heterarchical MPC architectures. As further development, it is till necessary to gain a clear understanding of the dynamics involved in the complex interaction of such a system and, in particular, it may be of interest to deeply investigate the behaviour of the proposed rule in a scenario in which the bottleneck resource change dynamically during the execution. Furthermore, in a semi-heterarchical architecture vision, it may be added an admittance system upstream the production system, in order to allow an additional balancing on the order to be admitted in the production system.
Figure 5: Production System Performances

Figure 6: Resources Utilization.
REFERENCES


**AUTHOR BIOGRAPHIES**

**GUIDO GUIZZI** is an assistant professor in the Department of Chemical, Materials and Production Engineering at the University of Naples Federico II. He has the title of Aggregate Professor in 2012. His current research interests include: modeling and simulation of stochastic systems, Supply Chain Management, and Operations Management, Logistics, Six Sigma. He obtained his PhD in Aerospace Engineering, Naval and Quality (University of Naples Federico II – Italy) in 2006. He has successfully collaborated with industry and academia, authoring more than 60 peer-reviewed research publications. He has been involved in several projects funded by companies. His email address is g.guizzi@unina.it.

**SILVESTRO VESPOLI** is a PhD Student in Technology, Innovation and Management at the University of Bergamo and the University of Naples Federico II, Italy. His research concerns the development of new methodologies and models for the design, management and control of production system in the framework of the digital factory (Industry 4.0) with the aim to support the design of sustainable processes and infrastructure, for both production and services of the industrialised county. The research grounds on strongly quantitative and interdisciplinary techniques, ranging from the use of control models, data science and forecasting techniques. Results of his research activity have been published in top-ranked international journals and in international conferences. His e-mail addresses are silvestro.vespoli@unina.it and silvestro.vespoli@unibg.it.

**ANDREA GRASSI** is Full Professor at the Department of Chemical, Materials and Production Engineering at the University of Naples Federico II, Italy. His current research interests are in the areas of design and management of manufacturing and logistics systems, including the development of stochastic analytical models, simulation modeling, performance analysis and prediction, Industry 4.0 distributed control architectures, and automation of manufacturing systems. Results of his research activity have been published in top-ranked international journals and in international conferences. He has been involved in several projects funded by companies and in European projects. His email address is andrea.grassi@unina.it.

**LIBERATINA CARMELA SANTILLO** is Full professor of Production Systems for Goods and Services at the Department of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II. Since 1998 – Professor of the following courses: Production Management, Operations Management, Industrial logistics at the undergraduate and graduate programs of Mechanical Engineering and Management Engineering at University of Naples Federico II. Her current research interests include modelling and simulation of stochastic systems, world-class manufacturing, and circular economy practices in manufacturing. She has successfully collaborated with industry and academia, authoring more than 120 peer-reviewed research publications. Her email address is santillo@unina.it.