Using online simulation in Holonic Manufacturing Systems
Olivier Cardin, Pierre Castagna

To cite this version:

HAL Id: hal-00620880
https://hal.archives-ouvertes.fr/hal-00620880
Submitted on 8 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Using online simulation in Holonic Manufacturing Systems

Olivier Cardin \textsuperscript{a,b}, Pierre Castagna \textsuperscript{a,b}

\textsuperscript{a} IRCCyN, 1 rue de la Noë, 44321 Nantes, France
\textsuperscript{b} IUT de Nantes, 2 avenue du Pr J. Rouxel, 44475 Carquefou, France

Corresponding Author: Olivier Cardin: olivier.cardin@irccyn.ec-nantes.fr
Tel: +33 (0) 228092025; Fax: +33 (0) 228092021

Abstract

This paper deals with the use of online simulation on Holonic Manufacturing Systems. Concepts needed for the use of online simulation in a classical hierarchical system were already defined, the observer being the central one. The behavior’s differences between both classes of systems are studied to determine the best way to adapt these concepts to this new environment. In the PROSA reference architecture, staff holons were chosen to welcome the simulation models and the observer. An application on an industrial sized Holonic Manufacturing System is described to demonstrate the validity of the approach.

Keywords:
Online simulation, HMS, Production Activity Control, PROSA, Staff holon.

1. Introduction

For several decades, classical hierarchical manufacturing architectures were used. In the past few years, distribution of information and control was explored until reaching purely distributed isoarchic architectures (Pujo and Ounnar, 2008). Among these new alternatives, Holonic Manufacturing Systems (HMS) now encounter a growing success in both academic and industrial worlds (Blanc,
Indeed, the combination between global fixed objectives imposed to the holarchy and local strategies left to the free will of these holons is a particularly attractive compromise between fully hierarchical and fully heterarchical behaviors.

The problem with such control architectures is their production activity control. Indeed, it is very difficult to mentally think through the consequences of alternative courses of action, and as a matter of fact to be able to determine whether an alternative decision is better than another. These difficulties, already pointed out in the 1990s about the only problem of scheduling (Van Dyke Parunak, 1991)(Cavalieri et al., 2007), are due to the distributed decision making, avoiding a global vision of the system’s behavior.

As shown in (Valckenaers et al., 2006), efficiency of a holonic architecture goes by the ability for holons to forecast the future behavior of the underlying system. An example of the importance of this problematic is indirectly presented in (Sousa and Ramos, 1999). When computing the scheduling of the HMS in a distributed manner, an indecision problem may occur during the negotiation between holons in charge of the resources (named Resource Holons) and the holons in charge of the orders (named Task Holons) on the only criterion of the due date. This indecision was solved by the authors by assigning a priority order in the negotiation depending on the date of arrival of the order, avoiding several concurrent negotiations on the same resource. This solution seems to be the best considering the holons cannot forecast the future behavior of the system.

One of the solutions suggested in (Valckenaers et al., 2006) to achieve this objective is to use ant colonies to predict the emergent behavior of the system on a short term (Valckenaers et al., 2003) (Valckenaers et al., 2006) (Karuna et al., 2005). This solution enables the order holons to gather information about its completion before it is launched (time of completion, date of arrival at each machine, etc.). Moreover, resource holons gather this information to build a projected planning of occupancy which may be used in the future negotiations with other holons.

The focus of this paper is to detail the advantages of using discrete-event simulation as an online forecasting tool. This possibility was already exposed in (Gouyon et al., 2007) when saying that the
use of tools supporting a discrete-event simulation of the production flows could ensure that no live-locks could occur when various products are considered at the same time in a product-driven system. Regarding their behavior, the underlying production systems can mainly be seen as discrete-event systems. As a matter of fact, discrete-event simulation is a very powerful tool to model HMS. Indeed, it is very suitable to describe with a lot of details the behavior of each holons in the holarchy. It may even be used as an emulation of the shop floor in order to study alternative rules applied to each holons before the real implementation (Blanc, 2006). For several decades, online simulation was described as a powerful tool for decision support in large and complex manufacturing systems. This simulation is meant to foresee the consequences of alternative decisions a decisional entities may take. Comparing the impacts, this entity is able to decide which solution is the best to apply. Unfortunately, this was almost never implemented as no general framework was designed to ensure the feasibility of the approach. It is only in the last years that (Cardin and Castagna, 2008) introduced several concepts which made the implementation of online simulation possible. This work particularly dealt with the initialization of online simulations, which was one of the biggest issue described in the literature, but only considered the case of complex hierarchical manufacturing systems.

The problem is to know whether this approach is compatible with the specificities of HMS. First, a short description of the PROSA architecture is exposed. This architecture was chosen among others for its balance between hierarchy and heterarchy. A model of decision making for both resource and order holons, based on cognitive engineering works, is exposed in section 2. These models clearly show that the decision making mechanisms are similar in both approaches, except for the localization of each step of the process. The third section deals with the initialization issue. Several propositions are made to efficiently gather the system’s state in order to initialize the simulations. This section introduces the use of Staff holons for decision support of other holons. This Staff holon includes both simulations and observer. This use is more precisely described in the last section, which introduces an example of online simulation on a flexible manufacturing cell.

2. Decision making in HMS
   
a. Holonic architectures
Many holonic architectures are exposed in the literature for the production activity control of HMS. One of the most famous, enabling a good compromise between hierarchy and heterarchy, is called PROSA (Van Brussels et al., 1998). As expressed in the acronym, PROSA (Product Resource Order Staff Approach) is mainly built around four kinds of holons (Figure 1).

First, the Product Holons (PH) deal with all the data related to the reference s produced in the HMS. They roughly act like a database for all the other holons of the system to enable them to access all the data (routings, bills of materials, etc.) needed to be able to produce the right products with a sufficient quality. The Resource Holons (RH) correspond to physical entities able to act on products (such as factories, machines or tools) and contain all the data and knowledge to be able to control and organize these entities. The Order Holons (OH) handle the product during its manufacturing, deal with the logistics needed in the routing of the product. These holons are meant to negotiate with RH and other OH in order to achieve at best the production. At last, the Staff Holons (SH) are used as specific advisors. When other holons (OH or RH) need to make a decision, they have to apply either the general fixed rules of the system, or a variant coming from the application of the local strategies. If this is not satisfactory, they may ask for the help of a particular SH in order to solve a specific problem. Each SH has therefore an expertise on very specific problems of the HMS. If the SH is not able to provide a result, or not soon enough, the other holons have the possibility to apply their rules. This results in a non-optimal but safe behavior of the HMS.
As a matter of fact, there are only two kinds of holons that are able to make decisions in a HMS: order holons and resource holons. It is to be noticed that the type of decisions they make are completely different. Indeed, RH are inclined to make decisions on a very local point of view, when OH have to cope with the global objectives of the system, and therefore make decisions not only on a local, but also on a global point of view. Furthermore, as every holons deal with fixed rules and local variants, the mechanisms of decision making are quite different.

b. Decision making model

(Hoc, 1996) presents a revised version (adding the cognitive mechanisms of situation evaluation) of Rasmussen’s model (Rasmussen, 1983) (Rasmussen, 1986) of human approach for problem solving in the diagnosis phase. Diagnosis is here defined as “a comprehension activity relevant to an action decision” (Cegarra and Hoc, 2001). The authors say that diagnosis is an activity of comprehension, organizing elements into a meaningful structure. This organization is oriented towards decisions relevant to actions. While diagnosing, the operator manages a balance between benefits and costs, trying to reach an acceptable performance according to the goals. The aim is to adapt this model, centered on the role of human operator, to the general case of HMS, centered on the holons.
Rasmussen’s model exposes three separate behaviors of the decisional entity (i.e. the holon). First, if it has already encountered the exact same situation, it enters a skill-based behavior (application of a simple rule: High Priority First for example). Otherwise, if the situation is close to a situation previously encountered, it enters a rule-based behavior (application of a well-defined procedure: use of a bin-packing algorithm for example (Blanc, 2006)). At last, if the situation has never been encountered before, or if the solving procedure is not well-defined, the decisional entity may enter a knowledge-based behavior, where a new solution, taking into account global objectives, is determined.

As stated in (Cardin, 2007), the use of online simulation in heterarchical architectures is best fitted in a rule-based behavior: skill based behavior generally does not need any simulation study, and knowledge-based behavior implies a model design that is not compatible with the HMS’s dynamic.

Simplified and adapted model of Rasmussen’s rule-based behavior is presented on figure 2.

Figure 2. Simplified Rasmussen’s rule-based behavior

c. Decision making using online simulation in a HMS
This model was designed by Rasmussen to study the role of human operators in scheduling. The operator is considered as a problem solver, when the holons generally have to negotiate inside the holarchy to obtain the same result. The decisional center is thus distributed (Bongaerts et al., 2000), which makes a great difference in the approach. Indeed, there may be several users of online simulation at a time in a HMS.

Decisions in a HMS are only taken in RH and OH - the holon making the decision will be called DH in the following. When this holon needs to make a decision, it asks for an advice to online simulation. This advice is meant to orient the decision towards the optimal one. The classical problem of such a behavior is the response time. Indeed, there is generally a time pressure on the DH to make its decision, so that it does not slow the production down. As a matter of fact, there is a possibility that online simulation is not able to provide the results early enough. In that case, the DH has to apply a default rule, known to be safe but not necessarily optimal.

One of the main hypotheses of this paper is to be able to model the decision making process in the simulation. Indeed, as the future behavior of the HMS depends on both the present and the forthcoming decisions, it is necessary to take into account the actual decision making process results to accurately predict this behavior. This hypothesis is not very restrictive as the simulation horizon is generally short in online simulations. This short horizon prevent from modeling events such as breakdowns for example, as their occurrence is too rare in the horizon. (Saint Germain et al., 2003) present a coupled real-time/as-fast-as-possible simulation, which enables a fast computation of the simulation with a shift to real time during the decision making process. It seems possible to use this approach if it is not possible to model the decisions. However, it has to be reconsidered if several replications have to be performed on stochastic models, as the simulation duration is quite longer.

**d. Implementation of online simulation in a HMS**

Online simulation is a tool designed to help the decision making in the PREVISION phase. Even if the decision is distributed, it is possible to centralize the simulation engines. As a matter of fact, the best place to implement online simulation is inside a staff holon. This holon is an aggregate of several
atomic RH, each of them corresponding to a simulation model devoted to a particular problem. The SH may also integrate optimization algorithms if the advice asked by the DH needs it.

Every holon have a remote access to run and collect the results of a simulation. This solution has the advantage of being easier to implement, but the model management is difficult (each specific request corresponding to a specific model) and it is necessary to multiply the simulation engine to allow multiple simulation at a time.

3. Initialization of online simulation

a. Introduction

However, the main problem identified in (Cardin and Castagna, 2008) still remains: the initialization of the simulation on the state of the HMS is both very important and very difficult to perform, as the data needed for this initialization are distributed among the holons. Three main possibilities of state gathering were examined: the following paragraphs present these possibilities. Their conclusions deal with the applicability conditions.

b. Data directly coming from the production system

The first solution studied was to obtain the data directly from the holarchy (Figure 3). Indeed, each holon has a partial knowledge of its state. For example, product holon may provide the production data, order holons may provide data concerning the production orders running, and the resource holon may provide the state of the work stations (breakdown, etc.). This solution provides obvious advantages: the obtained data are reliable, and represent the actual present state of the system. Furthermore, it only requires a minimal additional architecture to work properly.
Figure 3. Data directly coming from the production system

The most important component of this data collection is the performance of the sensors settled on the system. Some of them are able to report continuously the evolution of a variable: a temperature sensor may for example give the continuous conditions inside an oven. However, in general, a lot of data are not measured in this way. Let us consider for example the position of an Automated Guided Vehicle (AGV) inside a workshop. Sensors settled at certain points of the network (intersections, work stations, etc.) are sufficient to drive the fleet.

The remaining problem is that the use of such devices creates a great uncertainty. Indeed, when an AGV faces a sensor, it is possible to know its position at the current date. As soon as it moves between two of these sensors, its position is totally unknown (except for the information that it is situated on the track between the sensors), and its arrival date in front of the next sensor is unknown a priori. It is therefore not possible to have the whole fleet of AGV in front of sensors of the system each time the information of their location is needed. As a matter of fact, the holons only have a partial view of the system’s state: this is what was defined as the space and time uncertainties in (Cardin and Castagna, 2006).
Of course, one could argue that there is always a technical response to such a problem. In this work, the choice that was made is to use the only sensors strictly necessary to the control of the HMS. The previous example allows a better understanding of this hypothesis. The use of a GPS module, or a video camera coupled with an image analyzer is an efficient technical possibility to solve the problem of positioning of AGVs in a workshop. But, this type of solution is quickly expensive, and it would generally be necessary to add several of these technologies to know the complete set of data needed to build the entire state of the complete system.

The conclusion that was drawn from the use of this solution is that it is not applicable in the general case, as many data, essential in the definition of the state, are missing or unavailable through the sensors settled on the system. However, it could be recommended in the rare cases enabling its optimal use, as the data is particularly reliable considering the relative simplicity of the setup. A good example of application field is the Air Traffic Control (Rogers et al., 1991), where position, flight plan and current trajectory of each aircraft are perfectly known at any time.

c. Using a simulator

For the last few years, most of the simulation software editors provide “real-time” editions of their tools. These tools have the specificity to run at the same speed as the wall clock, in opposition to classical software in which the simulation engine runs as fast as possible. They were developed in order to meet the industrial requirements to be able to use simulation as an emulation of the production systems. The initial idea was to reduce the setup time of complex production systems by making tests of the control system directly on this emulation, without waiting for the existence of the physical system.

The idea here is to use such a simulation running in parallel with the production. This simulator is meant to react to production solicitations coming from the order holons in the same way the real system does. Doing so, it creates a permanent “image” of the system, which state can be considered as being the state of the real system at any time (Figure 4).
This solution has two main advantages compared to the exclusive use of data directly coming from the production system. First, it suppresses the space and time uncertainties problem that was defined in the previous section. About the example of the movement of an AGV previously presented, the simulator enables to know its current position at any time for example. Furthermore, the state gathered with this method is well formatted for a later computer treatment. Indeed, gathering the state is not useful if it is not possible to treat the data in an effective way. On the other hand, there are several problems in the application of this solution. First, the initialization of this simulator must be done with a perfectly known state. Typically, this state is chosen “idle and empty” because it is the easiest state to express by hand. Secondly, inaccuracies in the system’s model and possible hazards inevitably occurring along the production cause a deflection on the final results. The additive aspect of this deviation along time would end up having an effect far from being negligible.

To guarantee the applicability of this solution on a given production system, several constraints have to be set. First, the simulator has to be extremely reliable to limit the effects of modeling inaccuracies. Then, it is preferable to deal with systems which regularly go back to a known state (generally the “idle and empty” state mentioned before). Let us cite as a classical example the systems that are empty...
at the end of each working day. It is possible on such systems to restart the simulation, and by doing so, reset the deviation. Finally, the system shall have a low hazard occurrence frequency relatively to the length of simulation without reset. Indeed, these hazards are not taken into account into such a model, and are thus directly deviating the results.

As a matter of fact, the class of systems that may be concerned by this solution is relatively limited due to the inevitable growing deviation of the results. On the other hand, this is a credible alternative for applications that cannot afford the equipments necessary to the previous solution and which do not need a high precision on the results.

d. Using an observer

The conclusion that can be drawn at this point is that both solutions previously presented turned to be well adapted to a limited class of systems. As the objective of this work to make a solution applicable to the widest class possible, the idea is to get the advantages of the simulator (about the data availability) and of the direct use of real system’s data (about the data reliability). The following paragraph deals with a successful attempt of hybridization of the solutions in an observer.

Figure 5 presents the principles of such a solution. Compared to the previous solution, the simulator is replaced by an observer, meant to deal with the data coming from all the holons. Using these data, it is able to reset the deviation of its predictions. As a matter of fact, this observer is meant to run at the wall clock speed (real-time) as the simulator does.
This solution is probably one of the most powerful that can be implemented with the constraints presented before. But, even if the implementation costs are relatively low, this solution is not so easy to implement and is thus made for large facilities. The purpose of this paper dealing with the Holonic manufacturing systems, this constraint is very often respected as these systems are generally relatively large. A lot of technologies might be used to implement this observer, according to the objectives that were designed. The choice that was made here is to use discrete-event simulation. Indeed, a lot of simulation pieces of software meet the requirements of our study. First, it was widely used to model the behavior of such production systems. As a matter of fact, the model that was eventually made for the design of the facility can be used again in the production phase. This implies an interesting diminution of the investment time and costs. Then, the available means of communication are generally very well adapted to the communications inside such control architecture. Their graphical user interfaces is also very interesting, as it enables a clear vision on the behavior of the system for an operator. Finally, the state gathered on the observer is particularly well adapted to use to initialize online simulations in order to predict the future behavior of the system.

4. Example
The studied HMS (Figure 6) is located in the Quality, Industrial Logistics and Organization (QLIO) department of the University Institute of Technology (IUT) of Nantes, France. It is an automated assembly line, part of a larger set of logistics systems, including an AS/RS and two opto-guided AGV. This line is widely inspired from an actual line of FILTRAUTO/SOGEFI, world leader of the production of oil filters.

Figure 6. General overview of the system

This job-shop with automated transfers system is made of six workstations, each of them being an individual resource holon (RH) with its own intelligence. The goods are carried by a set of 40 transporters equipped with smart tags, all considered as separate RH. A transporter storehouse (RH) is available to store unused transporters. When a new order is placed, a new order holon (called main OH in the following) is created, made up of as many atomic OH as transporters involved in the production. This decomposition allows having two separate points of view on the same production order: global and local. Each atomic OH negotiates then to seize a transporter-RH, and lead it on the main loop,
where it can negotiate with each station RH in order to go on its recipe. When this recipe is over for all the goods the atomic OH was in charge of, it leads the transporter RH to the storehouse and free it.

As a matter of fact, to run an order, the main OH needs to make three decisions by negotiating with the other main OH (Figure 7 shows the localization of all these decisions on an assembly line scheme):

- DG1: At which date will the production begin?
- DG2: How many transporters will be allocated to this order?
- DG3: Which priority is given to the order?

The atomic OH also negotiate with the storehouse RH to decide:

- DG4: Does the transporter located on the main loop at the entrance enter the storehouse or stay on the loop?

Along the production, atomic OH and stations RH negotiate to decide:

- DL1: Does the transporter located on the main loop at the entrance enter the station or stay on the loop?
- DL2: At the end of an operation, is OH authorized to step to the next operation of the recipe?

Because of their physical structure, buffer management rule of stations 1, 3, 4 and 5 is necessarily FIFO. On the opposite, buffer management rule of stations 2 and 6 can be unspecified, which defines new buffer RH (Figure 8). As a matter of fact, two more decisions have to be negotiated between atomic OH, stations RH and the corresponding buffers RH:

- DL3: Does the atomic OH seize the station RH and free the buffer RH?
- DL4: Does the atomic OH free the buffer RH to enter the main loop?
Figure 7. The assembly line and the related decisions

Figure 8. PROSA architecture adapted to the assembly line

Obviously, all the decisions listed here do not need online simulation decision support. The following describes an example of application: DL3 decision. As stated before, this decision deals with stations 2
and 6 composed of a loop-shaped buffer with a programmable priority rule (Figure 9). The application is about the use of online simulation to choose which OH is supposed to enter the station RH.

![Diagram of Station 6](image)

**Figure 9. Implementation of station 6**

Numerous dynamic scheduling rules of the buffer can be used as a basis of DL3 decision. Most of these rules look for the best candidate among the elements waiting in the buffer, considering a given objective, to be taken out and treated by the station. The problem of these rules is that the buffer evolves along time, as the rules only take into account the content of the buffer at a precise date. However, it is not necessarily true that the choice made at date \(d\), taking into account the content of the buffer at this date \(d\), turns out to finally be the best solution. The idea here is thus to use online simulation to extend the vision of the rules to the near future (horizon \(H\)), to test the validity of this decision on this horizon.

(Cardin and Castagna, 2006) presented the use of the “Clear-a-fraction” rule applied to decision DL3. It is interesting to notice that the decision is totally automated, without any human intervention. This is due to the short decision lap of time left to the decisional holon (here a station RH). This constraint also has an implication on the use of online simulation. Indeed, to ensure a good working of the line, the station RH has to obtain the results before a due date. This leads to the use of a Staff Holon, which role would only be to advise the RH. If it cannot provide any result before the end of the due date, the station RH is meant to apply a simple rule (let the atomic OH enter the station for example).
Figure 10 describes the working of a simulation-based negotiation for the seizing of the station 6 RH by an atomic OH in a sequence diagram. After the request, the station checks the content of the corresponding buffer RH. To make the diagram easier to read, the alternative when the content of the buffer fulfill the conditions to directly grant the entrance to the OH is not represented. As a matter of fact, a simulation is requested to a Staff Holon with a timeout (called duedate in the diagram). This SH requests the state of the observer resource holon, and then computes the simulations. At this point, two alternatives: either the results are available on time and the station RH applies the recommendation of SH, or it is not provided on time and the station RH automatically grants the access.

![Sequence diagram of the seizing of station 6 RH](image)

5. Conclusion and future works

This paper deals with the use of online simulation on Holonic Manufacturing Systems. In the literature, all the concepts needed for the use of online simulation in a classical hierarchical system
were already defined (Cardin and Castagna, 2008). The simulation-made observer is the central concept, enabling the simulations to initialize on the actual present state of the system. In this classical approach, Manufacturing Execution System plays a central role, being the link between all the components of the architecture. In the PROSA reference architecture, this role is distributed among all the holons. As a matter of fact, the main aspects to study were both the place of online simulation in the architecture and the relationships between simulation and holons. Staff holons were chosen to welcome both the simulation models and the observer, advising the other holons in their decisions.

The final part of this paper is dedicated to the description of an application of these concepts to an industrial sized Holonic Manufacturing System. An example of fully automated buffer management is presented. This example allowed describing the behavior of each holons, in order to show the timed relationship between each of them. This specific case is in the class of goods routing problems, for which distributed architecture is very well adapted.

The future developments of this work will deal with the generalization of the approach to other control architecture. Furthermore, a real industrial application of the concepts to already existing manufacturing systems will enable a numbered comparison of performance between the use with and without online simulation.

6. References


