Allocating Non-Unitary Resources in a Constraint-Based Scheduler

G. Gini
F. Zaraga
Politecnico di Milano

M. C. Sioli
Zaraga Engineering, Milano

ABSTRACT: The new flexibility of industrial plants increases managerial and organizational problems. The number of alternatives increases, the mix of products grows, each product can be manufactured on different machines, the trend towards just-in-time production requires that the system react quickly, and reduces the temporal intervals for where to allocate activities. So the "only manual" approach to scheduling is inadequate. However, the present level of knowledge and expertise and the solutions so far used should be carefully considered. From this consideration we propose an architecture that could, in a smooth way, integrate the present solution and organization with new knowledge formulation and problem solving strategies.

We have studied the organization of an Italian company that produces shoe soles for the world market. The high mix of products and the flexibility required by the production facilities impact on the scheduling activity. Here, scheduling could be initially approached as a one-machine problem, where the machine is the curing press and has a non-unitary capacity. The allocation of this machine is governed by many constraints that account for a great variety of conditions, both technological and organizational.

We have defined the scheduling and resource allocation problem as a constraint satisfaction problem. We have provided a solution algorithm that orders the constraints and applies constraint checking. We have integrated the scheduling system with the existent PLC automation and with the central data and central processor. The result is a real-time system, running on a PC, and

International Journal of Expert Systems Pp. 481–498, Volume 9, Number 4, ISSN 0894-9077

producing the schedules for the next seven days. The scheduler is routinely called once a day, to take into account the data about real production. The production manager can call it in case of disrupting events.

Finally, we discuss the implementation and testing of our system.

1. INTRODUCTION

In this paper we approach scheduling with a focus on the requirements generated by manufacturing environments that are heavily subject to the effects of uncertainty and unpredictability (Meng & Sullivan, 1991; Paolucci et al., 1993; Smith et al., 1992). In this context, production plans must be frequently re-arranged to react to the occurrence of unexpected events. We based our work on investigations of a real productive context, the VIBRAM plant of Albizzate (Varese, Italy), where we examined the problem of short term production planning.

This approach is designed to support the decisional process of production managers and it constitutes an attempt to integrate organizational and mathematical issues into a general framework. The focus of attention about production scheduling in manufacturing is often directed toward the search for more efficient algorithms for solving such complex problems. As a result only the mathematical part of the problem is outlined, obtaining schedules that often are not satisfactory from a managerial and organizational point of view.

Our approach is designed to compute a feasible solution, able to sub-optimize the schedule about one or more criteria, as quickly as possible. This choice is based on two considerations:

- The ability to provide a good-enough solution in a short time is fundamental for managers in real manufacturing environments; and
- Mathematically speaking, a solution is not always guaranteed to exist. This situation cannot be accepted in a productive environment because the manager cannot wait for a long time to be notified that no solution exists. We need a system able to always compute a feasible solution.

The absence of solutions may depend on decisions made by the marketing department that has accepted too many orders, or on unexpected breakdowns of the system, or on personnel absence, or on raw materials unavailability. In any case, the production manager cannot lose the production. If a solution is not found, from a managerial viewpoint, the system is required to support users in "building" a satisfactory one. This approach, named cooperative scheduling, is presented in (Numao, 1994).

Another important aspect to consider is that unexpected events may simultaneously occur in different parts of the shop floor, discarding existing schedules and requiring fast rearrangements. The adjustment of an on-going schedule must be accomplished by managers of different departments.

The problem we have approached is the daily scheduling at VIBRAM, a producer of high-quality shoe soles made of various rubber materials. Here the due dates are given as the week number in the year, and the production is started in pull from the shipping department. Because of the highly personalized and fashion-sensitive production the inventory should be kept as small as possible (Fisher et al., 1994).

A serious scheduling problem arises in VIBRAM because of the high number of productive alternatives available. In fact, there are more than 18,000 articles organized into six classes (solid rubber, blown material, gumlite, two-colors, plates, eco-step) according to technological similarities; every article code indicates a size (there are up to ten sizes), a compound, and a color.

The production of shoe soles requires the work of three departments.

- First, the *dosing and mixing department* works the raw materials and prepares the compound that in a second pass is added with curing agents and divided into 'biscuits' of the desired weight and (possibly) shape;
- Second, the *curing press department* prints the soles on vulcanizing presses equipped with the correct molds and filled with biscuits; and
- Finally, the *finishing and packaging department* does the final cutting and polishing and a complete quality inspection. However, quality inspection is distributed between all the departments and mainly relies on people working on the machines.

The most crucial department is the second one, where the various molds should be loaded and filled to carry on the production. Several presses are organized into lines; each line and the relative team of workers are preferentially assigned to a few classes of production. Usually, the production mix is very high. and presses should be dynamically allocated to the products they can work.

Every article has one (or very few) mold(s), so for a given production we usually have a unitary resource (the mold) to be allocated on a non-unitary machine (the press that can simultaneously work on 4–10 molds), by means of a unitary resource (the worker that changes the molds). Each mold change takes 15–30 minutes, so in practice the number of changes in the two shifts of the day is limited (and is also agreed upon with the trade unions).

We have defined the scheduling problem as a constraint satisfaction problem (Fox, 1986), because the constraint approach is more flexible than the operational research approach and can manage a mix of objectives. We have integrated the scheduling system with the existent and grooving PLC automation, and with the company data base maintained on the central processor. The result is a real-time system, running on a PC network and producing the schedules for the next seven days. The scheduler is routinely called at the end of the day, to take into account the data about the real production. In case something disrupts the execution of the proposed plan, the system is called again by the production manager to produce a new schedule or to suggest alternative allocations.

2. SIZE OF THE PROBLEM AND OBJECTIVES

2.1. The Environment at VIBRAM

The production of shoe soles at VIBRAM requires the work of three production departments (for some products remotely located) as illustrated in Figure 1.

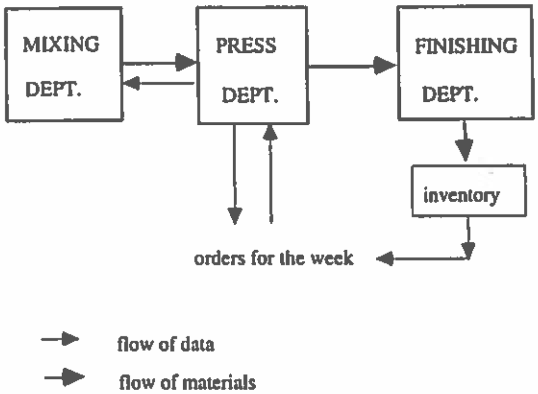

Figure 1. Organization of the Production Departments in VIBRAM

- First, the *dosing and mixing department* works in two steps. In the first step it doses and mixes the raw materials according to a locally defined schedule. With them it prepares the first compound, that can be stored for long times, and is produced on large capacity mixers. In the second pass the compound is added with curing agents and divided into 'biscuits' of the desired weight and (possibly) shape. The second compound can last only for one or a few days.
- The *curing press department* prints the soles on vulcanizing presses equipped with the correct molds (each containing left and right printings). Every article needs some settings for curing; usually time, temperature, and times of partial openings of the press characterize the controllable variables and define the cycle. The cycle is the same on all the presses of a line for the old electromechanical presses. For new presses equipped with PLC some local settings can be made. Moreover, some products need a lubricant agent to help the detachment from the mold, while others do not need it, and others would be damaged by the liquid.
- Finally, the *finishing and packaging department* does the final polishing and a complete quality inspection. However, quality inspection is distributed between all of the departments and relies only on people working on the machines.

The most crucial department to schedule is the second, where the various molds should be loaded on the machines to carry on the production. Several presses are organized into lines; each line is preferentially assigned a class of production according to the medium term production plan. Usually, the production mix is very high, and it is uncommon that two presses work the same article.

The main data about the scale of the problem of soles production in VIBRAM are:

- 18,000 articles in production (soles, half-soles, and plates);
- 1,200 clients;
- 24,000 orders in the year (each order contains more jobs defined by article/size/color);

- 10 to 5,000 size of a lot in production;
- The production is organized in two shifts a day;
- There are 5 lines, each with 8 to 16 presses; each press has 2 or 4 planes; each plane contains 3 to 5 molds. Each line can be operated only if a compatible team of workers is present; and
- 400 molds are in contemporaneous use.

2.2. The Schedule of the Press Department

In the existent organization, the production manager of the press department receives on Thursday morning a set of data about the orders to be filled for the next week, and another set about orders for the rest of the month. It is important to know whether the production of an article will last more than a week. In two days the manager is able to arrange the schedule for the next week and to send orders with due dates to the mixing department. Usually the schedule is reviewed at the end of every day, when production data are manually or automatically collected, and the approximate time of mold changes is set day by day.

The many objectives of the schedule include:

- To avoid order tardiness and unnecessarily early release;
- To maximize production;
- To respect due dates; and
- To coordinate with the mixing department to obtain in time the right quantity and size of biscuits.

2.3. The Data

Data are obtained from the data bases on the factory mainframe. They are:

- The orders data base;
- The articles data base;
- The mold data base; and
- The laboratory data base, that contains technological data about the articles, as temperatures and times of curing to get the optimal result. The curing time should be guaranteed to get a high-quality product. Longer times are acceptable, but should be avoided to maximize production.

During the introduction of PLC-based automation in the production department, the laboratory data have been made available also on the production computer. Data about the status of the machines, the number of pieces produced, and the number of defective soles are now available to the scheduler through the factory link.

Other sparse technological data are available only on paper. Most of the technological processing data were not available in any written form before the scheduler project was started; they have been collected from discussions with the production manager.

3. OUR APPROACH

Simple scheduling problems are found in allocating operations on a single machine. Our scheduler is much more complex, because the one-machine process is complicated by the availability of parallel machines, the need to consider the set-up time (time to load a new mold), the limited number of molds available, and the large number of molds that should fit into the non-unitary machine.

According to (Dorn, 1995) scheduling is the problem of sequencing a set of operations and allocating them to temporal slots without violating technical and capacity constraints. In production scheduling there is a set of resources (machines, workers, materials, . . .) and a set of jobs that require these resources. The processing of a job on a machine is called an operation. The problem is to allocate the resources to carry on the operations in a correct way.

Moreover, different correct solutions can be evaluated according to an evaluation function: the solution that minimizes the evaluation function is the optimal solution for the criterion expressed by the evaluation, called the objective. It is not easy to state the objectives in scheduling; more than 20 have been reported (French, 1982). So usually we refer to a few objectives that have practical relevance, for instance the minimization of the completion time, as we discuss in Section 5. Moreover, there is a trade-off between the minimization of one evaluation function and the growth of other performance measures: a schedule optimal for one objective can be very poor with respect to another one.

Many algorithms have been developed over the years to construct optimal schedules: they originate from Operations Research and use integer programming, dynamic programming, or approximation algorithms. They are all limited to the consideration of one goal function at a time, and to a small number of operations or machines. They do not manage realistic problems very well, where there is a mix of objectives and a large number of operations or resources. In practice, the most popular solution is the use of dispatching rules, that aim at reaching a goal under various conditions (French, 1982).

Since general scheduling problems are NP-complete (French, 1982), many authors prefer to find only a good solution using heuristics to prune the search space; this makes the solution difficult to generalize or to extend.

Scheduling and resource allocation are here viewed as constraint satisfaction problems (CSP) as in (Berry, 1992; Dahl, 1993; Keng & Yun, 1989; Zweben & Fox, 1994). The formulation of scheduling as a CSP deserves attention given the seminal work of (Fox, 1986).

A CSP is usually modeled by a set of variables {V1, ..., Vn}, with domains of possible values. {D1, . . . , Dn}, and a set of n-ary constraints among the variables {C1, . . . , Ck}. Finding the solutions for the network means finding all the combinations of values that satisfy all the constraints. Besides the obvious backtracking, many advanced algorithms (Kumar, 1992) have been developed for constraint satisfaction. Most of them rely on consistency checking algorithms, that propagate constraints to check node and path consistency, and reduce the constraint network. Sometimes the solution of the reduced problem is obvious. The limitation is that only unary and binary constraints are considered, while scheduling often requires n-ary constraints.

For instance, a variable in our scheduling problem is the height of the mold, that should fit the others on the plane; the domain of this variable is restricted by a set of binary equality constraints. Another variable is the size of the mold to allocate, whose domain is restricted by a n-ary constraint between the other molds sizes and the plane dimension.

Reducing the scheduling problems to a CSP does not mean reducing their NP intractability, since a CSP is still intractable. Moreover, the goal of the scheduling activity is to find an optimal, or at least a good schedule, not all the possible schedules. As a consequence, general methods for CSPs are still unable to solve the general scheduling problem but give ideas about how to make constraint propagation and constraint checking.

What are the advantages? The important points of this formulation are flexibility, expressiveness, and ease of modification. Constraint propagation, that is tractable, can be used for constructing the solution. To guide the selection of which variable to instantiate, and to which value in its domain, the use of AI techniques and a great deal of knowledge about the problem has demonstrated many important advantages (Zweben & Fox, 1994).

According to (Le Pape, 1994), AI techniques facilitate the expression and the application of factory-dependent knowledge together with theoretical knowledge. Moreover, the distinction between factual knowledge and control knowledge facilitates the adaptation of the system to different situations. The resulting schedule can be explained, the user can understand why a choice was made, and override it when necessary.

According to many authors, we have recognized and defined a large number of *hard* and *soft* constraints. Hard constraints should be satisfied, soft constraints are preference constraints, and can be relaxed. The constraints are explicitly defined or implicitly derived by other resource allocations.

3.1. Resources

Resources relevant for the scheduler are:

- *the presses*, with their splitting into planes;
- *the molds*, that usually consist of a unitary combination of an upper and a lower part. In two-color products they are a unitary combination of three parts; in other products upper and lower parts are less than the articles, and some (non-unitary) special combinations of them can work the different articles. Molds have different (discrete values) heights and amplitude.
- *the molds changer*, who is responsible for taking the molds from the inventory, changing it, and returning the molds to the inventory.

A curing press is a very complex resource. Allocating a job (a mold) to a press is critical because of many hard constraints. Consider that each press has a few planes; on each plane, that has a given dimension, we want to put the maximum number of molds. Usually 4 molds (for 8 printings) can be arranged, but many exceptions arise for large sizes. Moreover, the height of the molds on the same plane should be the same, the lubrication agent the same, and the working cycle (temperature, time of openings, curing time, number and time of partial openings, etc.) should be the same for all the planes of a press.

3.2. Constraints

The main problem in scheduling is that many constraints should be satisfied, and usually they are conflicting. The manager considers the following constraints:

- *compatibility constraints* (geometrical and technological) that should be always respected. Geometrical constraints are about the size and the height of the molds that can be accommodated on the same plane. Usually the big numbers have big molds, and the heuristic used by the managers is to group big sizes on the same plane. Technological constraints are about the temperature and time of curing that depend on the mixing, the size, and the color. Other technological constraints are about the need for lubricant agents. The continuous values of the controlled variables are usually translated into discrete values by the chemical laboratory that makes the initial tests and gives discrete values according to a table. The production manager has some rules to be able to modify some of the technological constraints to accommodate production under different conditions.
- *organizational constraints* are about the "best" use of workers, the line balancing, the preferred time for some operations, etc. Examples are using the plane as a LIFO to reduce the labor for the mold changer, mixing on the same plane molds that require a short loading time (one or few pieces of rubber) with only one that requires a long loading time to maximize the speed of the line, etc. They are usually preferential constraints. Another rule usually applied is to print all the biscuits (usually produced in some excess). The motivation is that the number of defective soles is determined only at the end of the production process, when the biscuit can be too old and the mold already taken away; so it is better to produce some pairs in excess. This rule introduces uncertainty in the real size of jobs, because the exact number of the biscuits is unknown, and it would be too expensive to count them in advance.
- *time constraints* are simple with respect to other kinds of production. The main time constraint is about reducing waste: when a second compound arrives in the presses area, it can last for a short time, so that compound should be printed. Another time constraint is the time needed to load the press, that should be the same for each press in the whole line because presses open up at given time intervals and only one worker is charged with loading them in a fixed order. This time constraint is partially reduced to a set of organizational constraints. The most critical time constraint is about the mold changer: he needs a minimum time interval between two changes.

3.3. General Criteria and Heuristics

We aim to avoid the need to backtrack through intelligent decision making. Our scheduler is used both to generate a complete schedule (as it happens for Saturdays) and to insert new jobs into an existing schedule (as it happens every day).

Let's consider the insertion of new jobs: at a given moment presses are working with maximum capacity. In an order-oriented perspective each order to be filled is waiting for a press. When an order has finished production, the scheduler should choose a new order. The new order can start if its mold and biscuits are available, and if it is compatible with the rest of the orders of work on the plane, and at least partially compatible with other orders in the same press.

Compatibility is expressed by the degree of satisfaction of a set of hierarchical constraints: the higher the level of the violated constraint, the lower the degree of compatibility. The system sorts the relevant constraints and checks them:

- the geometrical constraints (space on the plane and same height) are respected;
- the technological constraints are compatible (or require a simple local modification of the settings);
- the due date is respected;
- the preferential constraints about the position in the plane are eventually satisfied;
- the number of manual operations on the involved line is in the average; and
- finally, the change of the molds are limited to the maximum set by the management.

To be sure that all the jobs have a chance to be scheduled, we assign a priority to the jobs. We start examining the orders (and so the molds they require) from the ones with highest priority. The priority is re-computed each time the system is run. The priority combines a value inversely grooving with the remaining time to deliver the order, and a value defining the importance of that order as given by the management.

If the system finds a job completely compatible with the finished one, it allocates it. Otherwise, it starts looking for partially compatible orders (relaxing in hierarchical order a soft constraint). In this way constraint satisfaction is reduced to simple constraint propagation.

Totally incompatible orders are allocated only when their priority is very high. Preferentially this allocation is done for Monday, according to a heuristic given by the management. In fact, on Saturday, the scheduler is run in an extended version to find the best organization of the molds for the coming week. On Saturday, since production is suspended, the maintenance service does the polishing of molds and can completely reorganize the molds and add new ones in large amounts without the production loss that characterizes mold changes during work days.

The preferential allocation of lines of presses to different products is done by the production management. For this reason we can start considering in a separate and hierarchical way the allocation of the different classes of jobs. Only to fill up unused positions, or to accelerate a late job, presses can be dynamically allocated to other classes of production. So, in our resource-oriented reasoning, the allocation on the lines to the jobs is done in a hierarchical order.

First, the production of the solid rubber articles that have a very short vulcanizing time is allocated, to fill the fastest lines. Usually, solid rubber articles can be grouped into three classes with similar technological data. The presses with the shortest settled cycle are loaded first. In general, three lines with fast cycle time are reserved.

The blown material articles that require longer curing times are usually allocated on two lines. Note that for those articles the settings of times and temperature is very critical; as a result of small variations a part of the obtained production may be smaller than the required dimensions. For these products the allocation should go from bigger to smaller sizes, so the already produced pieces that are out of measure can fill part of the order not yet allocated. Some critical products in this line can be allocated only in some positions of the planes to avoid damaging to themselves or to other products.

The same criteria apply to the other classes of products (gumlite, eco-step, two-colors, plates). In the near future the scheduler will be extended to schedule other classes of products that are assigned to sub-contractors.

The maximum number of lines and of sub-contractors to allocate to each kind of production is decided by the management, according to the production plan, and is fixed for a week, at least.

In Figure 2 we resume the basic line of reasoning followed by the scheduler.

At time t_1 the order o_i will terminate. The scheduler has to find which order to allocate in the position of the plane that will be free. It scans the list of the pending orders, and applies the constraints to find the one with the best compatibility, represented as an integer number from 1 to 8. For instance, #1 means that the new order has the same article/size/color of the finishing one, #2 means that it has the same mold with different color, and going up in the numbers we have lower degrees of compatibility. In the last levels of compatibility we have the high priority orders that require heavy rearrangement of the plane. After a level of compatibility for the order is found, constraint checking stops for that order. The system allocates the order with the best compatibility and computes its ending time.

The user input indicated in Figure 2 is a way to express the criterion to maximize and is optional. In practice it gives a different order to the levels of compatibility of the orders.

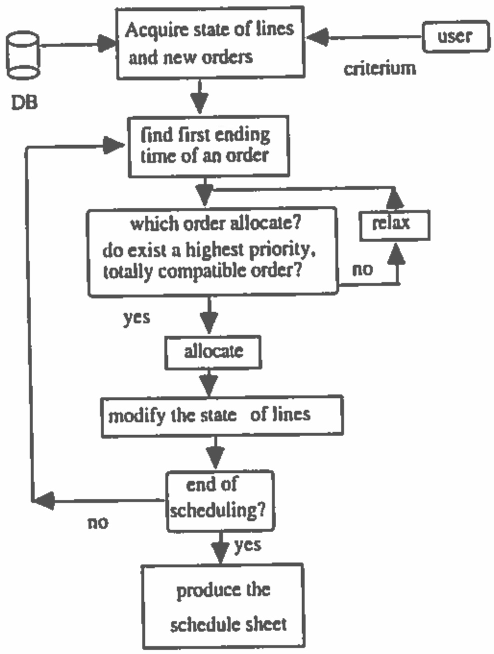

Figure 2. The Scheduler

Since those levels depend also on the priority and on the size of the order, changing their numbering has the effect of using different allocation policies. They may be relevant to assess the quality of the schedule, as discussed in Section 5.

3.4. The Scheduler and the Basic Automation

Before the introduction of our scheduler the press department was "manually" organized. A corporation program allocated the orders on the shop floor for the next week considering the mean production capacity. It was unable to produce the molds allocation and it did not know anything about the real production problems.

The production manager made the daily schedules for the next week and produced orders for the mixing department. The planned activity was changed only to allocate very important orders or to finish the biscuits produced in excess. The curing-press department was governed by a daily schedule sheet, containing a list of all the new articles to be produced on the lines and a list of color changes and of the mold changes, with approximate time. It was impossible to have complete information about the work in progress and the status of the lines.

The expert system has evolved with the introduction of PLC-based automation on some new lines, and a production monitor. For this reason it can access, at any moment, the complete information about the status of the presses, the orders, and the molds. The scheduler still needs to wait for data only from the old presses.

4. IMPLEMENTATION

The scheduler runs on a 486-based PC under Windows 3.1, connected to the mainframe computer and the factory link. The software has been written and compiled in ART-IM and ART-Enterprise, a shell of the Inference Co. (Los Angeles) that supports object-oriented programming and production rules, and in C language. This choice is made for efficiency reasons, because general constraint satisfaction languages are still inappropriate. Other authors are trying to compile rules from constraint solutions, we have manually produced object representations and rules.

The knowledge is represented in ART as classes, facts, and rules.

In an object-oriented representation we use classes that inherit properties from their superclasses. This is a useful way to represent the machines: lines of presses, presses, planes, and the molds allocated on the planes. Orders and articles are also represented as classes. For instance, the following is the definition of the order class:

```
(defschema ORDER
   (article code )
   (number of pairs 0)
   (time 0)
   (number of needed molds 0)
```

```
  (molds already allocated 0)
  (status of the order 0))
```

Facts have the role of storing temporary data in the working memory.

Rules express how and when to apply a constraint; their effect is to reduce the set of orders under consideration until an order to allocate is found. In the following example we see a simplified version of the rule that states the total compatibility of an order with a finished one:

```
(defrule TOTAL-COMPATIBILITY
;;; if an order is terminated and another order contains the same article
 ?a (examined-order ?ordine-in&~NIL ?comp-vulc ?comp)
(not (selected-criterion ok))
(schema ?plan
  (free-space ?dimensione-libera)
  (plan-code ?codice-piano))
(schema ?position
  (is-a ?plan)
  (state-of-position FREE)
  (order-in-position ?order-out)
  (due-date NIL)
  (ending-time NIL)
  (mold-in-use ?stampo-in-uso))
(schema ?order-in
  (instance-of order)
  (art-col-mis ?art-col-mis))
(schema ?order-out
  (instance-of order)
  (art-col-mis ?art-col-mis))
(schema selection-criteria
  (criterion1 ?criterio 1 &: (eq ?criterio 1 0)))
⇒
;;; then mark this order as good to allocate
(retract ?a)
(bind ?lista (CREATE$ ?ordine-in ?comp-vulc ?comp ))
(modify
  (schema selection-criteria
```

```
        (criterion 1 ?lista)))
    (free$ ?lista) )
```

The hierarchical approach is implemented using the priority mechanism provided by ART.

Programs in C are used to access the data bases, and to read from the factory link data about the status of the shop floor; they produce data files for the expert system. Another C program produces and prints the sheet of the mold changes, according to the results of the expert system.

Different functions are accomplished by different units, implemented on different machines as illustrated in Figure 3:

- The long term production plan is generated on the host;
- The daily and weekly production schedules are produced on the Department system;
- The real production is monitored by the production monitors; and
- Presses are activated and production data sent to the monitor at the control level. Meaningful data are the number of pieces produced and their code. To get the real number, defective parts and unused press cycles should be subtracted. Monitoring is automatic for the presses controlled with PLC, manually organized for the other presses.

The manager through the production monitor can check the state of the presses and their real settings and productivity. Since he can access knowledge not available in the system (for instance knowledge about the skills of available personnel), he can input managerial data (for instance the decision to stop production, etc.) that the scheduler cannot produce.

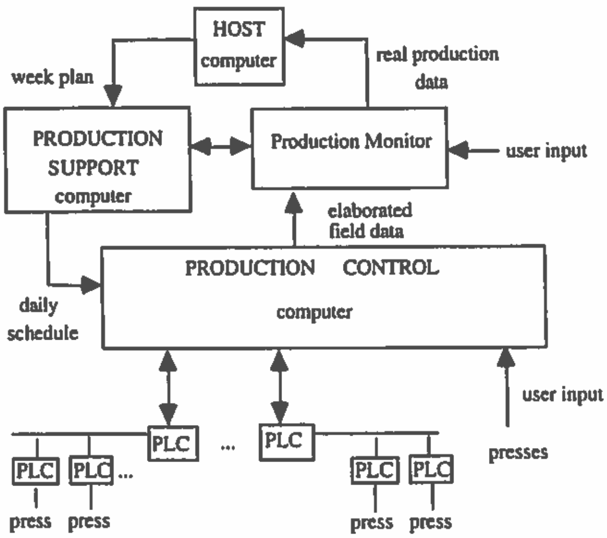

Figure 3. Functional Architecture of the Production System

The scheduler can react to events known at the production monitor as:

- A press is out of work;
- A mold needs maintenance, is broken or out of order for the present schedule;
- No material is available to load the press;
- A line is closed (not enough personnel in the shift); and/or
- An order with very high priority arrives.

The program produces a schedule for one week; every day we get a new schedule for the next 7 working days. The schedule is produced in a few minutes.

In 1994 (Gini, 1994) the scheduler was implemented on the production computer in a test version; the connections with the mainframe computer and the PLC automation were simulated. The testing and the results of the scheduling system in VIBRAM will be discussed in Section 6.

5. SCHEDULE QUALITY

Scheduling is very closely tied with production costs and organizational aspects; every schedule is supposed to deal with the minimization of costs related to the following drivers:

- Global duration of manufacturing process (makespan);
- Work in progress (WIP) costs;
- Equipment idle periods; and
- Tardiness costs.

The total cost of a schedule is given both by tangible and intangible elements (tardiness cost is particularly tied with the intangible aspects of making a brand appreciated). These parameters are connected to one another; there exists a trade-off (French, 1982) between the minimization of one of them (which can be in a particular situation considered as the most relevant) and the growth of the others.

Schedule quality represents the extent to which a schedule meets the required performance (by mixing different cost drivers) and the extent to which an organization is able to respect it. Usually schedules are obtained through the minimization of a parameter and the search for satisfactory values for the others, disregarding the real requirements of organizations that have to accomplish them (Muscettola, 1993).

Manufacturing organizations are interested in generating high quality schedules in order to minimize their production costs and to have feasible schedules. To get such schedule quality the solution methods should:

1. Provide a good solution in a sufficiently short time; it also has to be able to recognize if a feasible solution exists; and
2. Evaluate the effects of changing constraints and priorities on the problem solution. In fact, by changing the value of some constraints, it's possible to improve schedule quality. This requirement is particularly felt in distributed manufacturing environ-

ments: each production managers would like to know what is the effect that their decision (concerning for example orders splitting or orders priorities) have on final schedule quality, on other organizational units, and on global productive efficiency. This is what happens in our case when negotiating with the mixing department.

Although these two capabilities are independent, both are indispensable in achieving schedule quality. According to this analysis our work was directed toward the definition of a support to managerial work, able to reduce the response time and to evaluate managers proposals.

In the application to the VIBRAM press department it came out that the most important driver is tardiness cost, because the equipment was never idle and WIP could not be reduced too much because of the limited number of molds that impose to work in parallel on different orders.

We have experimented with the system with different criteria aiming at maximizing the efficiency of a solution with respect to any of the illustrated drivers. The manager could generate schedules under those different criteria and choose the best according to the value of the relevant criterion. After experimentation, it came out that managers preferred the schedule that reduces the average tardiness, and this has been set as the default choice.

However, to allow changing this choice, we have introduced in the user interface a new parameter, R (reactivity to late orders), whose value is a real number in the range 0÷4.9. Before the introduction of the ES, this parameter was very near to 0 (low reactivity), while now the default choice is 4.

6. EFFECTS OF THE SCHEDULER

VIBRAM has a tradition in using advanced software technologies. They are long time users of CAD, CAM, and CASE systems. However, the introduction of the expert system was followed with some uneasiness.

The first version of the scheduler was installed at VIBRAM at the beginning of 1994. It was reduced to only a subset of the articles, and was based on a set of rules collected and approved from the managers.

During the field testing, it was run for two months and the results compared with the manually produced ones. The results indicated that the number of mold changes was higher than in the manual schedule, and that not all the available biscuits were printed. The system did not know the number of biscuits, so it set a mold change after the completion of an order, while the practice of producing much more compound than needed was common. During this time the production manager spent a lot of time in making the manual schedule and checking it with the output of the expert system.

In the meantime the system was debugged, extended, and provided with the support software in C.

During Spring 1994 the system was installed in the extended version and used, for a limited number of lines, by the manager who did not produce the schedule anymore.

After a few months a system able to manage all the lines and the products was available. Extensions are always under way, for instance to connect with the mixing department.

Table 1

	before ES	*with ES*
R	0	4
# of mold charges	1	9
# of cycles lost for mold changes	5	45
# of pairs lately produced	240	45
# of late orders	8	0

After one year of experimentation some results have been achieved. At the level of the factory management, the main positive effect is a sensible reduction of late orders. At the production level, the positive effects are many. The schedule is automatically produced and available at the beginning of the production shift, also when managers are not at work. The production manager has more time to follow other activities of the department. The negative effect of wasting the remaining biscuits has been reduced in two ways: the biscuits are ordered by the system in a more reasonable quantity, and the manager can ask the worker responsible for a line to finish the production of all the available material.

The reduction of the late orders had, after a while, the positive effect of reducing the number of mold changes, that was inflated by the implicitly stated rule of delaying (to maximize the productivity) the small orders that were always late.

The use of the system to reorganize the molds on Saturday is able to reduce the cycle times for some articles and the loss of production for the mold changes.

Quantitative results are more difficult to obtain, in part because no more manual schedules are produced, in part because it is not easy to determine what improvements are due to the expert system, and what to the introduction of factory automation. However we made some experiments on the values of R. Before the introduction of the ES, R was always near to 0, while 4 is now the default. To understand its meaning in production we can look at some data about a week of scheduling in the gumlite lines.

We ran the expert system twice: in the first trial we set R to 0, to simulate the choices usually made by the human schedulers; in the second we aimed at eliminating lateness by reactivity, so we set R to 4. We got the results illustrated in Table 1.

As expected, a more reactive system is able to reduce the lateness at the cost of slightly reducing productivity. According to our other tests we have assessed that the number of mold

Table 2

	before ES	*with ES*
% of late orders	35	5
% of pairs lately produced	60	10
# of weeks of lateness for one order	8	2
minutes for a mold change	120	15

Table 3

	before ES	*with ES*
# of pairs produced	100	100
# of pairs in the right week	40	90
# of pairs produced for next weeks	60	10
# of weeks of earliness	8	3

changes and the number of late orders are inversely correlated. We base our claim on values obtained during the testing phase of the ES, resumed in Table 2.

Another important value to consider is the number of orders released too early and going to the inventory. Table 3 shows those values on 100 pairs.

The scheduler is considered a valid help. It generates correct schedules, and can react to some unexpected events. It cannot be used as a mathematical tool to minimize any evaluation function, but the knowledge in it guides the system to generate satisfactory solutions.

7. CONCLUSIONS

We have approached the problem of scheduling the presses and mold allocations in a curing-press department looking at the strategies already employed in manual operations. We have formalized those strategies as a huge set of constraints. The human strategy gave us an idea of the hierarchy of constraints; so we were able to reduce the complexity of the search because constraints are applied in a given order. The philosophy of allocation is mainly order-based, because we start computing priorities of the jobs. A resource-based allocation is also considered because we avoid unused positions to attain maximum productivity. The filling of those positions is sometimes done by grouping together difficult jobs, or taking low-priority jobs.

The need for a scheduling system was found by the management as a consequence of the increased flexibility and level of basic automation of the plant. The production manager was almost entirely working on scheduling. Our idea to mimic human reasoning resulted in a system accepted by the management, moderately diverging from the manual solutions in the number of mold changes. This number should be low to increase productivity, but needs to be high enough to reduce tardiness. A computer system is better than humans in finding all the instances where a change can be done.

At present the extension of the scheduling system to integrate the dosing and mixing department is being tested. The solution is based on negotiation, starting with a request from the press department. The problem is that the mixers have very high capacity, so the compounds cannot be prepared in the small quantity often needed but it is necessary to group orders that need the same compound. This is an instance of an optimization criterion that can make more convenient the operation of the mixing department, but it is in contrast with the operation criterion of the press department. For this reason a negotiation step is considered. Moreover, the second compound cannot be stored, so it is necessary to allocate it to presses, even in the case when the schedule of the press department is different. On the other end, the integration of the finishing department is not needed. In fact the work there is based on human labor and humans have a high degree of flexibility.

Acknowledgments: Partial support for this research has been obtained through the Special Project on Planning of the Italian CNR. Thanks are given to the anonymous referees for their valuable and constructive suggestions and comments.

REFERENCES

Berry, P.M. (1992, Spring). A predictive model for satisfying conflicting objectives in scheduling problems. *AI Magazine*.

Dahl, V. (1993). Guest Editor's introduction: Constraint-reasoning for expert systems. *Int. J. of Expert Systems, Research and Applications*, *6*, 4.

Dorn, J. (1995, March). Iterative improvement methods for knowledge-based scheduling. *AICOM*, *8*, 1.

Fisher, M.L., Hammond, J.H., Obermeyer, W.R., & Raman, A. (1994, May–June). Making supply meet demand in an uncertain world. *Harvard Business Review*.

Fox, M. (1986). Observations on the role of constraints in problem solving. *Proceedings of Sixth Canadian Conference on A.I.*, Montreal.

French, S. (1982). *Sequencing and scheduling: An introduction to the mathematics of the job-shop*. Chichester, UK: Ellis Horwood Ltd.

Gini, G., Sioli, M.C., & Zaraga, F. (1994). A system for allocating a complex resource in manufacturing. *SIGMAN workshop at AAAI 94*, Seattle.

Keng, N., & Yun, D. (1989). A planning/scheduling methodology for the constrained resource problem. *Proc. 11th IJCAI*, pp 998–1003, Detroit.

Kumar, V. (1992, Spring). Algorithms for constraint-satisfaction problems: A survey. *AI Magazine*.

Le Pape, C. (1994). Scheduling as intelligent control of decision-making and constraint propagation. In M. Zweben, & M. Fox (Eds.), *Intelligent scheduling*. San Francisco, CA: Morgan Kaufmann.

Meng, C.C., & Sullivan, M. (1991, February). Logos—A constraint-directed reasoning shell for operations management. *IEEE Expert*, 20–28.

Muscettola N. (1993). An experimental analysis of bottleneck-centered opportunistic scheduling. *Workshop W20* at IJCAI 93, Chambery, France.

Numao, M. (1994). Development of a cooperative scheduling system for the steel-making process. In M. Zweben, & M. Fox (Eds.), *Intelligent scheduling*. San Francisco, CA: Morgan Kaufmann.

Paolucci, E.M., Preda, S., & Gini, G. (1993). Scheduling in SOS. *Workshop W20* at IJCAI 93, Chambery, France.

Smith, S.F., Keng, N., & Kempf, K.G. (1992). exploiting local flexibility during execution of pre-computed schedules. In A. Famili, D. Nau, & S. Kim (Eds.), *AI Applications in Manufacturing*. Menlo Park, CA: AAAI Press/MIT Press.

Zweben, M., & Fox, M.S. (1994). *Intelligent scheduling*. San Francisco, CA: Morgan Freeman.

Address all correspondence to: G. Gini, Dept. of Electronic Engineering and Information Sciences, Politecnico di Milano, Italy.