Ant colonies for performance optimization of multi-components systems subject to random failures

Nabil Nahas, Mustapha Nourelfath and Daoud Ait-Kadi
Universite Laval, Mechanical Engineering Department CIRRELT
Canada

1. Introduction

Reliability has been considered as an important design measure in industry. The design of a system involves in general numerous discrete choices among available component types based on reliability, cost, performance, weight, etc. If the design objective is to maximize reliability for a certain cost requirement, then a strategy is required to identify the optimal combination of components and/or design configuration. This leads to combinatorial optimization problems which are NP-hard. For large industrial problems, exact methods are lacking since they require a very large amount of computation time to obtain the solution of the problem. This chapter will focus on the use of ant colonies to solve three optimal design problems which are among the most important in practice:

1. The reliability optimization of series systems with multiple-choice constraints incorporated at each subsystem, to maximize the system reliability subject to the system budget. This is a nonlinear binary integer programming problem and characterized as an NP-hard problem.

2. The redundancy allocation problem (RAP) of binary series-parallel systems. This is a well known NP-hard problem which involves the selection of elements and redundancy levels to maximize system reliability given various system-level constraints. As telecommunications and internet protocol networks, manufacturing and power systems are becoming more and more complex, while requiring short developments schedules and very high reliability, it is becoming increasingly important to develop efficient solutions to the RAP.

3. Buffers and machines selections in unreliable series-parallel production lines: we consider a series-parallel manufacturing production line, where redundant machines and in-process buffers are included to achieve a greater production rate. The objective is to maximize production rate subject to a total cost constraint. Machines and buffers are chosen from a list of products available in the market. The buffers are characterized by their cost and size. The machines are characterized by their cost, failure rate, repair rate and processing time. To estimate series-parallel production line performance, an analytical decomposition-type approximation is proposed. Simulation results show that this approximate technique is very accurate. The optimal design problem is formulated

as a combinatorial optimization one where the decision variables are buffers and types of machines, as well as the number of redundant machines.

For each problem, a literature review will be presented. This review will focus on metaheuristics in general, and on ant colonies in particular. Recent advances on efficient solution approaches based on Hybrid Ant Colony Optimization (HACO) will be detailed.

2. Reliability optimization of a series system

2.1. Literature review

As it is often desired to consider the practical design issue of handling a variety of different component types, this paper considers a reliability optimization problem with multiple-choice constraints incorporated. To deal with such reliability optimization problems with multiple-choice constraints incorporated, Sung and Cho [9] have used an efficient branch-and-bound method. Nourelfath and Nahas [6] have solved the reliability optimization problem by using quantized neural networks. [11] deals with a reliability optimization problem for a series system with multiple choice constraints incorporated to maximize the system reliability subject to the system budget. The problem is formulated as a binary integer programming problem with a non linear objective function [1], which is equivalent to a knapsack problem with multiple-choice constraints, so that it is NP-hard [3]. Some branch-and-bound methods for such knapsack problems with multiple-choice constraints have been suggested in the literature [5,7,8]. However, for large industrial problems, these methods are lacking since they require a very large amount of computation time to obtain the solution of the problem. This section describes the use of an *ant system* to obtain optimal or nearly optimal solutions very quickly.

2.2. Optimal design problem

Let us consider a series system of n components. For each component, there are different technologies available with varying costs, reliabilities, weights and other characteristics. The design problem we propose to study is to select the best combination of technologies to maximize reliability given cost. Only one technology will be adopted for each component. In order to formulate the problem in mathematical expression, the following notations are introduced first:

n the number of components
M_i the number of technologies available for the component i
C_i^j the cost of a component i using the technology j (C_i^j is assumed to be known)
R_i^j the reliability of the component i when the technology j is used
B the available budget
TC the total cost.

We specify the decision variable x_i^j (with $j = 1, 2, \ldots, M$; and $i = 1, 2, \ldots, n$) as:

$$x_i^j = \begin{cases} 1 \text{ if the component } i \text{ uses the technology } j \\ 0 \text{ otherwise} \end{cases}$$

Considering these notations, the proposed series-system reliability optimization problem is expressed in the following binary nonlinear integer programming problem:

Maximize

$$Z = \prod_{i=1}^{n} \left(\sum_{j=1}^{M_i} x_i^j R_i^j \right)$$

Subject to

$$\sum_{i=1}^{n} \sum_{j=1}^{M_i} x_i^j C_i^j \leq B \tag{1}$$

$$\sum_{j=1}^{M_i} x_i^j = 1 \ \forall \ i = 1, 2, \ldots, n \tag{2}$$

$$x_i^j = \{0, 1\} \ \forall \ j = 1, 2, \ldots, M_j \text{ and } i = 1, 2, \ldots, n \tag{3}$$

Constraint (1) represents the budget constraint; without loss of generality, we consider that B is an integer. Constraint (2) represents the multiple-choice constraint, and constraint (3) defines the decision variables.

When a solution satisfies all the constraints, it is called a feasible solution; otherwise, the solution is said to be infeasible. Our goal is to find an optimal solution or sometimes a nearly optimal solution. This is motivated by the fact that in real size industrial systems, the search space of the reliability optimization problem formulated in this paper is very large, taking the use on non heuristic approaches infeasible. Ant system is a recent kind of meta-heuristic which has been shown to be suitable (especially when combined with local search) for combinatorial optimization problems with a good neighborhood structure (see e.g. [6,10]), as in the case of the reliability optimization problem formulated in this paper.

2.3. Solution approach of the reliability optimization problem

To apply the ant system (AS) algorithm to a combinatorial optimization problem, it is convenient to represent the problem by a graph $G = (\varsigma, \Lambda)$, where ς are the nodes and Λ is the set of edges. To represent our problem as such a graph, the set of nodes g is given by components and technologies, and edges connect each component to its available technologies. Ants cooperate by using indirect form of communication mediated by pheromone they deposit on the edges of the graph G while building solutions.

Informally, our algorithm works as follows: m ants are initially positioned on the node representing the first component. Each ant will construct one possible structure of the entire system. In fact, each ant builds a feasible solution (called a tour) by repeatedly applying a stochastic greedy rule, called, the *state transition rule*. Once all ants have terminated their tour, the following steps are performed:

- An improvement procedure is applied. This procedure, which will be detailed later, is composed of a specific *improvement algorithm* (called algorithm 1) and a local search.
- The amount of pheromone is modified by applying the *global updating rule*.

Ants are guided, in building their tours, by both heuristic information and by pheromone information. Naturally, an edge with a high amount of pheromone is a very desirable choice. The pheromone updating rules are designed so that they tend to give more pheromone to edges which should be visited by ants.

The state transition rule used by the ant system is given in equation (4). This represents the probability with which ant k selects a technology j for component i:

$$p_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{m=1}^{M_i} [\tau_{im}(t)]^\alpha \cdot [\eta_{im}]^\beta} \tag{4}$$

where τ_{ij} and η_{ij} are respectively the pheromone intensity and the heuristic information between component i and technology j . α is the relative importance of the trail and β is the relative importance of the heuristic information η_{ij}. The problem specific heuristic j information used is $\eta_{ij} = \dfrac{R_i^j}{C_i^j}$ where R_i^j and C_i^j represent the associated reliability and cost. That is, technologies with higher reliability and smaller cost have greater probability to be chosen.

During the construction process, no guarantee is given that an ant will construct a feasible solution which obeys the budget constraint (i.e. $\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{M_i} x_i^j C_i^j \leq B$). The unfeasibility of solutions is treated in the pheromone update: the amount of pheromone deposited by an ant is set to a high value if the generated solution is feasible and to a low value if it is infeasible. These values are dependent of the solution quality. Infeasibilities can then be handled by assigning penalties proportional to the amount of budget violations. In the case of feasible solutions, an additional penalty proportional to the obtained solution is introduced to improve its quality.

Following the above remarks, the trail intensity is updated as follows:

$$\Delta \tau_{ij} = \sum_{k=1}^{m} \Delta \tau_{ij}^k \tag{5}$$

ρ is a coefficient such that (1 - ρ) represents the evaporation of trail and $\Delta\tau_{ij}$ is :

$$\tau_{ij}(new) = \rho\tau_{ij}(old) + \Delta\tau_{ij} \tag{6}$$

where m is the number of ants and $\Delta\tau_{ij}^k$ is given by:

$$\Delta\tau_{ij}^k = \begin{cases} Q.penalty_k & \text{if the k}^{th}\text{ ant chooses technology } j \text{ for component } i \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where Q is a positive number, and $penalty_k$ is defined as follows:

$$penalty_k = \begin{cases} \left(\dfrac{B}{TC_k}\right)^a & \text{if } B < TC_k \\ \left(\dfrac{R_k}{R^*}\right)^b & \text{if } B \geq TC_k \end{cases} \tag{8}$$

B is the available budget, TCk is the total cost obtained by ant k, R_k is the reliability obtained by ant k and R^* is the best obtained solution. Parameters a and b represent the relative importance of penalties. It can be easily seen from the above equations that by introducing a penalty function, we aim at encouraging the AS algorithm to explore the feasible region and infeasible region that is near the border of feasible area and discouraging, but allowing, search further into infeasible region.

It is well known that the performance of AS algorithms can be greatly enhanced when coupled to local improvement procedures [2]. Following this, two local improvement algorithms are included in our AS approach (called local search algorithm and algorithm 1). Algorithm 1 uses the remaining budget (the amount not used by the ant) to improve the solution. In fact, some generated feasible solutions do not use the entire available budget.

This algorithm improves the initial solution by using this remaining budget to exchange some actual technologies by more reliable other technologies. A similar idea can be found in [10] where a neural network is presented to solve the job-shop scheduling problem, and where a similar procedure is used to improve the obtained solutions by eliminating the time segments during which all machines are idle.

The local search algorithm proceeds to change in turn each pair of chosen technologies by another pair. For each component, technologies are indexed in ascending order in accordance with their reliability. A solution $S = \{u, v, ...\}$ indicates that component 1 uses technology with index u, component 2 uses technology with index v, etc. Let consider for example a series system with 3 components and 6 available technologies for each component. Suppose that the obtained solution at a given cycle is $S = \{3, 2, 5\}$. The local search will evaluate the following solutions:

$$S=\{4, 1, 5\}, S = \{4, 2, 4\}, S = \{2, 3, 5\}, S = \{2, 2, 6\}, S = \{3, 3, 4\}, S = \{3, 1, 6\}.$$

Among all these evaluated solutions, whenever an improvement feasible solution is detected, the new solution replaces the old one. It has been shown in [11] that the experimental results showed that the optimal or nearly optimal solutions could be obtained quickly. In the next section, Hybrid Ant Colony Optimization (HACO) will be used to solve the redundancy allocation problem. This HACO uses rather the extended great deluge algorithm as a local search within the proposed ant colony algorithm.

3. Redundancy allocation problem

3.1. Problem description

The redundancy allocation problem (RAP) is a well known combinatorial optimization problem where the design goal is achieved by discrete choices made from elements available on the market. The system consists of n components in series. For each component i ($i = 1, 2, ..., n$) there are various versions of elements, which are proposed by the suppliers on the market. Elements are characterized by their cost and weight according to their version. Each component i contains a number of elements connected in parallel. Different elements can be placed in parallel. A component i is functioning properly if at least k_i of its p_i elements are operational (k-out-of-n: G).

The series-parallel system is a logic-diagram representation for many design problems encountered in industrial systems. As it is pointed out in [18] and [31], electronics industry is an example where the RAP is very important. In fact, in this industry most systems require very high reliability and the products are usually assembled and designed using off-the-shelf elements (capacitors, transistors, microcontrollers, etc.) with known characteristics. Other examples where the above type of structure is becoming increasingly important include telecommunications systems and power systems. In all these systems, redundancy is indeed a necessity to reach the required levels of reliability and the RAP studied in this paper is therefore one of the major problems inherent to optimal design of reliable systems.

Assumptions
1. Elements and the system may experience only two possible states: good and failed.
2. The system weight and cost are linear combinations of element weight and cost.
3. The element attributes (reliability, cost and weight) are known and deterministic.
4. Failed elements do not damage the system, and are not repaired.

5. All redundancy is active: failure rates of elements when not in use are the same as when in use.
6. The supply of elements is unlimited.
7. Failures of individual elements are s-independent.

Notation

R_{sys} overall reliability of the series-parallel system
R^* optimal solution
C cost constraint
W weight constraint
n number of components
i index for components
a_i number of available elements choices (i.e., versions) for component i
r_{ij} reliability of element j available for component i
w_{ij} weight of element j available for component i
c_{ij} cost of element j available for component i
x_{ij} number of element j used in component i
$\mathbf{x}_i$ $\left(x_{i1}, \ x_{i2}, \ ..., \ x_{ia_i} \right)$
p_i total number of elements used in component i
p_{max} maximum number of elements in parallel
k_i minimum number of elements in parallel required for component i
$\mathbf{k}$ $(k_1, k_2, ..., k_n)$
$R_i(\mathbf{x}_i|k_i)$ reliability of component i, given k_i
$C_i(\mathbf{x}_i)$ total cost of component i
$W_i(\mathbf{x}_i)$ total cost of component i

The RAP is formulated to maximize system reliability given restrictions on system cost and weight. That is,

Maximize
$$R_{sys} = \prod_{i=1}^{n} R_i (\mathbf{x}_i|k_i) \tag{9}$$

Subject to
$$\sum_{i=1}^{n} C_i (\mathbf{x}_i) \le C, \tag{10}$$

$$\sum_{i=1}^{n} W_i (\mathbf{x}_i) \le W. \tag{11}$$

Constraints (10) and (11) represent respectively the budget and the weight constraints. If there is a pre-selected maximum number of elements which are allowed in parallel, the following constraint (12) is added:

$$k_i \le p_i \le p_{\max} \quad \forall \ i = 1, \ 2, \ ..., \ n. \tag{12}$$

3.2. Literature review

The RAP is NP-hard [17] and has previously been solved using many different optimization approaches and for different formulations as summarized in [39], and more recently in [32]. Optimization approaches to determine optimal or very good solutions for the RAP include dynamic programming, e.g. [12,25,35,41], mixed-integer and nonlinear programming, e.g.

[32], and integer programming, e.g. [14,27,28,34]. Nevertheless, these methods are limited to series-parallel structures where the elements used in parallel are identical. This constitutes a drawback since in practice many systems designs use different elements performing the same function, to reach high reliability level [31]. For example, as explained in [18], (airplanes use a primary electronic gyroscope and a secondary mechanical gyroscope working in *parallel*, and most new automobiles have a redundant (spare) tire with different size and weight characteristics forming a 4-out-of-5: G standby redundant system). Because of the above-mentioned drawback and of the exponential increase in search space with problem size, heuristics have become a popular alternative to exact methods. Meta-heuristics, in particular, offer flexibility and a practical way to solve large instances of the relaxed RAP where different elements can be placed in parallel.

Genetic algorithm (GA) is a well-known meta-heuristic used to solve combinatorial reliability optimization problems [18,33,37,42,43]. In addition to genetic algorithms, other heuristic or meta-heuristic approaches have also been efficiently used to deal with system reliability problems. A tabu search (TS) meta-heuristic [30] has been developed in [31] to efficiently solving the RAP, while the ant colony optimization meta-heuristic [20] is used in [4] to solve the problem.

In light of the aforementioned approaches, the method presented here gives a heuristic approach to solve the RAP. This method combines an *ant colony* optimization approach and a degraded *ceiling local* search technique. This approach is said to be *hybrid* and will be called *ACO/DC* (for Ant Colony Optimization and Degraded Ceiling).

The idea of employing a colony of cooperating agents to solve combinatorial optimization problems was recently proposed in [21]. The ant colony approach has been successfully applied to the classical traveling salesman problem [22,23], to the quadratic assignment problem [26], and to scheduling [13]. Ant colony shows very good results in each applied area. The ant colony has also been adapted with success to other combinatorial optimization problems such as the vehicle routing problem [15], telecommunication networks management [24], graph coloring [19], constraint satisfaction [38] and Hamiltonian graphs [44]. In [11], the authors used ant system to solve the optimal design problem of series system under budget constraints. The ant colony approach has also been applied for the RAP of multi-state systems in [36]. For the RAP in the binary state case, which is the focus of the present paper, the only existing work is that of [4].

3. 3. Hybrid solution approach: ACO/DC

As for the problem studied in section 2, to apply the AGO meta-heuristic to this problem, it is convenient to represent the problem by a graph $G = (\varsigma, \Lambda)$, where ς are the nodes and Λ is the set of edges. To represent our problem as such a graph, we introduce the following sets of nodes and edges [55]:

- *Three sets of nodes*:
    1. The first set of nodes (N_1) represents the components.
    2. The second set of nodes (N_2) represents, for each component, the numbers of elements which can be used in parallel. For example, if the maximum number of elements in parallel is three ($p_{max} = 3$), the set N_2 will be given by three nodes corresponding to one element, two parallel elements and three parallel elements.
    3. The third set (N_3) represents the versions of elements available for each component.

- *Two sets of edges*:
  1. The first set of edges is used to connect each component node in the set N_1 to the corresponding nodes in N_2.
  2. The second set of edges is used to connect the nodes in N_2 to the nodes in N_3 of their available versions.

Informally, our algorithm works as follows: m ants are initially positioned on a node representing a component. Each ant represents one possible structure of the entire system. This entire structure is defined by the vectors $\mathbf{x}_i$ (i = 1, ..., n). Each ant builds a feasible solution (called a tour) to our problem by repeatedly applying stochastic greedy rules (i.e., the *state transition rules*). Once all ants have terminated their tour, the amount of pheromone on edges is modified by applying the *global updating rule*. Ants are guided, in building their tours, by both heuristic information (they prefer to choose "less expansive" edges), and by pheromone information. Once an ant has built a structure, the obtained solution is improved by using a local search algorithm. This step is performed only in the following cases:

- the obtained solution by the ant is feasible and,
- the quality of the solution is "good". The term "good" means here that the reliability R_{sys} of the structure should be either better than the best solution R^*, i.e., $R_{sys} \geq R^*$, or close to this best solution R^*, i.e., $0 \leq \dfrac{R^*-R_{sys}}{R^*} \leq +l$ where l represents the solution quality level.

Ants can be guided, in building their tours, by pheromone information and heuristic information. Naturally, an edge with a high amount of pheromone is a very desirable choice. The pheromone updating rules are designed so that they tend to give more pheromone to edges which should be visited by ants.

In the following we discuss the state transition rules and the global updating rules.

State transition rules

In the AGO algorithm, each ant builds a solution (called a tour) to our problem by repeatedly applying two different *state transition rules*. At each step of the construction process, ants use: (1) pheromone trails (denoted by τ_{ij}) to select the number of elements connected in parallel and the versions of elements; (2) and a problem- specific heuristic information (denoted by η_{ij}) related to the versions of elements.

An ant positioned on node i (i.e. component i) chooses the total number p_i of elements to be connected in parallel. This choice is done by applying the rule given by:

$$P_{ip_i} = \begin{cases} \dfrac{\left(\tau_{ip_i}\right)^{\alpha_1}}{\displaystyle\sum_{k=1}^{p_{max}} \left(\tau_{ik}\right)^{\alpha_1}} & if \quad p_i \in \{1, \ 2, \ .., p_{max}\} \\ \\ 0 & otherwise \end{cases} \tag{13}$$

where α_1 is a parameter that controls the relative weight of the pheromone (τ_{ip_i}). We favour the choice of edges which are weighted with greater amount of pheromone.

When an ant is positioned on node p_i representing the selected number of elements connected in parallel in component i, it has to choose these p_i versions of elements. This choice is done by applying the rule given by:

Ant colonies for performance optimization
of multi-components systems subject to random failures
483

$$P_{p_i j} = \begin{cases} \dfrac{\left(\tau_{p_i j}\right)^{\alpha_2} \left(\eta_{p_i j}\right)^{\beta}}{\sum\limits_{k=1}^{a_i} \left(\tau_{p_i k}\right)^{\alpha_2} \left(\eta_{p_i k}\right)^{\beta}} & if \quad j \in \{1, \ 2, \ .., \ a_i\} \\ 0 \quad otherwise \end{cases}$$ (14)

where α_2 and β are respectively parameters that control the relative weight of the pheromone ($\tau_{p_i j}$) and the local heuristic ($\eta_{p_i j}$).

We tested different heuristic information and the most efficient was $\eta_{p_i j} = (r_{ij})/(w_{ij}^3)$ where r_{ij} and w_{ij} represent respectively the associated reliability and weight of version j for component i. In equation (14) we multiply the pheromone on edges by the corresponding heuristic information. In this way we favour the choice of edges which are weighted with smaller weight and greater reliability and which have a greater amount of pheromone.

Global updating rule

During the construction process, no guarantee is given that an ant will construct a feasible solution which obeys the constraints (11) and (12). The unfeasibility of solutions is treated in the pheromone update: the amount of pheromone deposited by an ant is set to a high value if the generated solution is feasible and to a low value if it is infeasible. These values are dependent of the solution quality. Infeasibilities can then be handled by assigning penalties proportional to the amount of cost and weight violations. Thus, the trail intensity is updated as follows:

$$\tau_{ij}(new) = \rho \tau_{ij}(old) + \Delta \tau_{ij}$$ (15)

ρ is a coefficient such that (1 - ρ) represents the evaporation of trail and $\Delta \tau_{ij}$ is :

$$\Delta \tau_{ij} = \sum_{k=1}^{m} \Delta \tau_{ij}^k$$ (16)

where m is the number of ants. Furthermore, $\Delta \tau_{ij}^k$ is given by:

$$\Delta \tau_{ij}^k = \begin{cases} Q \ penalty_k \ R_{sys}^k & \text{if the edge } (i, j) \text{ is visited by the } k^{th} \text{ ant} \\ 0 & \text{otherwise} \end{cases}$$ (17)

where Q is a positive number, R_{sys}^k is the system reliability for ant k, and *penalty_k* is defined as follows:

$$penalty_k = \begin{cases} \left(\dfrac{C}{TC_k}\right)^a & if \ (C < TC_k) \\ \left(\dfrac{W}{TW_k}\right)^b & if \ (W < TW_k) \\ \left(\dfrac{C}{TC_k}\right)^a \left(\dfrac{W}{TW_k}\right)^b & if \ (C < TC_k \ and \ W < TW_k) \end{cases}$$ (18)

and TW_k are respectively the total cost and the total weight obtained by ant k. Parameters α and β represent the relative importance of penalties.

The degraded ceiling local search meta-heuristic

The principle of the degraded ceiling meta-heuristic

The performance of algorithms based on the AGO meta-heuristic can be greatly enhanced when coupled to local improvement procedures. A degraded ceiling (DC) based algorithm is included in our AGO approach to improve the solutions obtained by the ants.

The degraded ceiling is a local search meta-heuristic recently introduced in [6a] and [6b]. Like other local search methods, the degraded ceiling iteratively repeats the replacement of a current solution s by a new one $s*$, until some stopping condition has been satisfied. The new solution is selected from a neighbourhood $N(s)$. The mechanism of accepting or rejecting the candidate solution from the neighbourhood is different of other methods. In degraded ceiling approach, the algorithm accepts every solution whose objective function is more or equal (for the maximization problems) to the upper limit L, which is monotonically increased during the search by ΔL.

The initial value of ceiling (L) is equal to the initial cost function $f(s)$ and only one input parameter ΔL has to be specified. In [16], the authors applied successfully the degraded ceiling on exam timetabling problem and demonstrated that it outperformed well-known best results found by others meta-heuristics, such as simulated annealing and tabu search. They showed two main properties of the degraded ceiling algorithm:

- The search follows the degrading of the ceiling. Fluctuations are visible only at the beginning, but later, all intermediate solutions lie close to a linear line.
- When a current solution reaches the value where any further improvement is impossible, the search rapidly converges. The search procedure can then be terminated at this moment.

The degraded ceiling algorithm is an extension of the "great deluge" method which was introduced as an alternative to simulated annealing. Degraded ceiling, simulated annealing and "great deluge" algorithms share the characteristic that they may both accept worse candidate solutions than the current one. The difference is in the acceptance criterion of worse solutions. The simulated annealing method accepts configurations which deteriorate the objective function only with a certain probability. The degraded ceiling algorithm incorporates both the worse solution acceptance (of the "great deluge" algorithm) if the solution fitness is less than or equal to some given upper limit L, i.e. ($f(s*) \geq L$), and the well-known hill climbing rule ($f(s*) \geq f(s)$).

Like simulated annealing, the degraded ceiling algorithm may accept worse candidate solutions during its run. The introduction of the dynamic parameter has an important effect on the search. As explained in [16], the decreasing of L may be seen as a control process, which drives the search towards a desirable solution. Note finally that degraded ceiling algorithm has the advantage to require only one parameter (ΔL) to be tuned.

3.4. Test problems and results

The test problems, used to evaluate the performance of the ACO/DC methodology for the RAP, were originally proposed by Fyffe *et al*. in [25] and modified by Nakagawa and Miyazaki in [35]. Fyffe *et al*. [25] specified constraint limits of 130 units of system cost, 170 units of system weight and $k_i = 1$ (i.e., l-out-of-n:G). Nakagawa and Miyazaki [35] developed 33 variations of the original problem, where the cost constraint C is set to 130 units and the

weight constraint *W* varies from 159 units to 191 units. The system is designed with 14 components. For each component, there are three or four element choices [55].

Earlier optimization approaches to the problem (e.g., [25] and [35]), required that only identical elements be placed in redundancy. Such approaches determined optimal solutions through dynamic programming and IP models, but only a restricted set of solutions was considered due to computational or formulation limitations of exact solution methods. Nevertheless, for the ACO/DC approach, as in [81], [31] and [4], different types are allowed to reside in parallel. In [18], Coit and Smith first solved the RAP with a genetic algorithm without restricting the search space. More recently, Kulturel-Konak *et al.* solved this problem in [31] with a tabu search algorithm, while Liang and Smith [31] used an ant colony optimization approach improved by a local search. Because the heuristic benchmarks for the RAP where elements mixing is allowed are the methods in [18], [31] and [4], there are chosen for comparison. Our approach will be compared also with [35] and [29]. By comparing the proposed ACO/DC methodology to all the above-mentioned papers (e.g., [18], [31], [4], [35] and [29]), we compare it to the best-known solutions found in literature at the best of our knowledge.

In meta-heuristics such as AGO, simulated annealing and degraded ceiling, it is necessary to tune a number of parameters to have good performance. The user-specified parameters of the ACO/DC algorithm were varied to establish the values most beneficial to the optimization process. Following the tuning procedure used in [21-23], we tested various values for each parameter, while keeping the others constant. Based on these initial experiments the values found to be most appropriate are:

$$\alpha_1 = 0.1, \alpha_2 = 0.5, \beta = 1, Q = 0.01, \rho = 0.9, a = 1, b = 10, \tau_0 = 1, \Delta L = 0.0001 \text{ and } l = 0.01.$$

These parameters values are used for all test problems. 50 ants are used in each iteration. When combined to the degraded ceiling algorithm, AGO converges quickly to optimal or near optimal solutions. Note that the degraded ceiling is called only if the obtained solutions are very good. For the considered problem instances, the maximum number of iterations needed does not exceed 300 iterations.

Comparing the results obtained by our approach with those of previous works [18,29,31,4,35], it has been shown in [55] that:

1. The solutions found by our algorithm are all better than those of Hsieh [19].
2. In 31 of the 33 test cases, the solutions found by our algorithm are better than those of Nakagawa and Miyazaki [27] while the rest (i.e., 2 cases) are as good as those they found.
3. Cases 22 to 29 and 31 to 32 are as good as those found by the genetic algorithm of Coit and Smith [8] while the rest (i.e., 24 instances) are all better than those they found.
4. In 6 of the 33 test cases, the ACO/DC outperformed the tabu search algorithm of Kulturel-Konak *et al.* [21] while it was very close but at a lower reliability in only 2 instances.
5. In 9 of the 33 test cases, the solutions found by our algorithm are better than those of Liang and Smith [24] while the rest are as good as those they found.

Both the degraded ceiling and the ant colony algorithms are meta-heuristics. Our contribution is based on the ACO/DC hybridization and very good results are obtained. The RAP is one of the most difficult combinatorial optimization problems inherent to optimal design of reliable systems. We believe and we show that two efficient meta-

heuristics have to cooperate in order to solve efficiently this problem, namely the AGO and the DC meta-heuristics.

The study conducted in this section shows again that hybridization of meta-heuristics is a very promising methodology for NP-hard reliability design problems.

4. Selecting machines and buffers in series-parallel production lines

4.1. Optimal design problem

Consider the series-parallel production line. Buffers are inserted to limit the propagation of disruptions, and this increases the average production rate of the line. This line consists of n components and n-1 buffers. Each component of type i (i = 1,2,..., n) can contain several identical machines connected in parallel. For each component i, there are a number of machine versions available in the market.

In order to formulate the problem in mathematical expression, the following notations are introduced first:

h_i version number of machine i
H_i maximum h_i available
$\mathbf{h}$ $\{ h_i \}, h_i \in \{1,2,.., H_i \}$
r_i number of elements connected in parallel in i
R_i maximum r_i allowed
$C(h_i)$ cost of each machine of version h_i
$P(h_i)$ isolated production rate of machine with version h_i
$T(h_i)$ processing time of machine with version h_i
$\lambda(h_i)$ failure rate of each machine with version h_i
$\mu(h_i)$ repair rate of each machine with version h_i
$u(h_i)$ speed of the machine's version h_i

We assume that a buffer is also chosen from a list of available buffers. The buffers of different versions f differ by their size $N^b{}_f$ and cost $C^b{}_f$. The total number of different buffer versions available for m^{th} component is denoted by F_m. The vector $\mathbf{f} = \{f_m\}$, where $0 \leq f_m \leq F_m$, defines versions of buffers chosen for each component. The entire production line structure is defined by the vectors $\mathbf{h}$, $\mathbf{r}$ and $\mathbf{f} = \{f_m\}$.

For the given $\mathbf{h}$, $\mathbf{r}$ and $\mathbf{f}$, the total cost of the production line can be calculated as:

$$C_T = \sum_{m=1}^{n-1} C_{f(m)}^{B} + \sum_{i=1}^{n} r_i . C(h_i)$$

(19)

The optimal design problem of production system can be formulated as follows: find the system configuration $\mathbf{f}$, $\mathbf{h}$ and $\mathbf{r}$ that maximizes the total production rate P_T such that the total cost C_T is less or equal to a specified value C^*. That is,

Maximize $P_T (\mathbf{f}, \mathbf{h}, \mathbf{r})$ (20)

Subject $C_T (\mathbf{f}, \mathbf{h}, \mathbf{r}) \leq C^*$ (21)

The input of this problem is C^* and the outputs are the optimal production rate $P_{T\text{Max}}$ and the corresponding configuration determined by the vectors $\mathbf{f}$, $\mathbf{h}$, $\mathbf{r}$. The resulting maximum value of P_T is written $P_{T\text{Max}} (C^*)$.

4.2. Literature review

There is a substantial literature on the analysis of production lines with buffers [45]. This literature is mainly concerned with the prediction of performance. Much of it is aimed at evaluating the average production rate (throughput) of a system with specified machines and buffers. In [46], the authors present a set of algorithms that select the minimal buffer space in a flow line to achieve a specified production rate. The algorithms are based on analytical approximations of the Buzacott model of a production line [47,48]. For a recent review of the literature on production line optimization, the reader is referred to [46]. The goal of the existing works is to choose buffers sizes for a production line. They all assume that the number of machines is specified, and the only parameters to find are buffers sizes. The proposed approach to optimal design aims at *selecting both buffers and machines*; it gives also the number of redundant machines used in parallel.

To deal with the optimal design problem considered in this work, it is mandatory to develop a method for throughput evaluation of series-parallel manufacturing production lines. This method has to take into account two characteristics:

(i) Components may consist of banks of parallel machines. Concerning this first characteristic, we attempt to represent each stage by an equivalent single component.

(ii) The processing rate may differ from component to component. To deal with this second characteristic, we use a continuous (or fluid) material model type which produces very good results. This consists of two main steps. First, the non homogeneous line is transformed into an approximately equivalent homogeneous one. In a second step, the resulting homogeneous line is analysed by using the well-known decomposition method for homogeneous lines [49].

The effect of the used simplifications for estimating throughput is examined by comparing the results provided by our approximate technique to simulation results on many examples. This comparison shows that the proposed approximate technique is very accurate.

As the formulated problem is a complicated combinatorial optimization one, an exhaustive examination of all possible solutions is not realistic, considering reasonable time limitations. Because of this, we develop two heuristics to solve the problem. The first heuristic is inspired from the *ant colony system* meta-heuristic: a recent kind of biologically inspired algorithms [48,49]. The second proposed heuristic is based on the *simulated annealing* meta-heuristic [50].

4.3. Throughput evaluation of series-parallel production lines

4.3.1. Summary of the method

The proposed method approximates each component (i.e. each set of parallel machines) of the original production line as a single unreliable machine. The system is then reduced to a single machine per component production line of the type represented in figure 2. The equivalent machines may have different processing rates. To determine the steady state behaviour of this *non-homogeneous* production line, it is first transformed into an approximately equivalent homogeneous line. Then, the well-known Dallery-David-Xie algorithm (DDX) is used to solve the decomposition equations of the resulting (homogenous) line [51].

4.3.2. Replacing each component by an equivalent machine

The decomposition techniques developed in the literature are efficient in estimating performance characteristics of series production lines. In these techniques it is necessary for

each component to be described by three parameters: *failure rate*, *repair rate* and *processing rate*. By limiting the description of the equivalent machine to these three parameters, our analysis of the new system is reduced in complexity to that of the existing decomposition techniques. Furthermore, the state space of a series-parallel line grows large with the number of parallel machines in the components. Replacing each set of parallel machines by one equivalent machine leads advantageously to a reduction of the state space.

Let denote by λ_{ij}, μ_{ij} and P_{ij}, respectively, the failure rate, the repair rate and the processing rate of a machine M_{ij}, and by λ_i, μ_i and P_i, respectively, the failure rate, the repair rate and the processing rate of a machine M_i. To calculate the three unknown quantities λ_i, μ_i and P_i, we have to formulate three equations. Assuming that machines within the set of parallel machines are fairly similar, the following approximation is proposed in [52] :

$$\lambda_i = \sum_{j=1}^{J_i} \lambda_{ij} \qquad\qquad i = 1, 2, \ldots, n \qquad\qquad (22)$$

$$\mu_i = \sum_{j=1}^{J_i} \mu_{ij} \qquad\qquad i = 1, 2, \ldots, n \qquad\qquad (23)$$

$$P_i = \sum_{j=1}^{J_i} P_{ij} \qquad\qquad i = 1, 2, \ldots, n \qquad\qquad (24)$$

It is shown in [52] that it is a good approximation. However, when buffers are small, this heuristic is inaccurate. In the present work, we assume that the available buffers are large enough. Thus, each set of parallel machines is approximated as an equivalent single machine by using equations (22), (23) and (24). This leads to a non-homogenous line. Therefore, a homogenisation technique is required, as explained in the next subsection.

4.3.3. Homogenisation technique

It is known that in the case of non-homogenous lines (*i.e.* production lines in which the machines do not have the same processing time), two approaches can be used. The first approach is based on an extension to the case of homogenous lines of the decomposition technique developed in [49]. The second approach relies on the modification of the non-homogeneous line into an approximately equivalent homogeneous line by means of *homogenisation techniques* [53]. The analysis of the obtained homogeneous line is therefore based on the use of the decomposition method for homogeneous line. In this way, it is possible to rely on the DDX algorithm which has been proven to be very fast and reliable. In [53], the authors showed that the homogenization method of [54], referred to as the completion time approximation, provides fairly accurate results. In this method, each machine M_i of the original non-homogeneous line is replaced by an approximately equivalent machine M_i^e, such that its completion time distribution is close to that of the original machine. The processing time of machine M_i^e is set to the processing time of the fastest machine in the original line: $T^e = \min (T_1, T_2, \ldots, T_k)$. Since the processing time of M_i^e is given (equal to T^e) there are two parameters per machine that must be determined, namely the failure and repair rates. Let λ_i^e and μ_i^e be the failure and repair rates of machine M_i^e. The principle of the method developed in [53] is to determine λ_i^e and μ_i^e in such a way that the distribution of completion times of machine M_i^e has the same first and second moments as

those of the distribution of completion times of machine M_i. The values of λ_i^e and μ_i^e are given in [53] by:

$$\lambda_i^e = \left[\frac{T_i}{T^e}\left(1+\frac{\lambda_i}{\mu_i}\right)-1\right]^2 \frac{T^e}{T_i}\frac{\mu_i^2}{\lambda_i} \text{ and } \mu_i^e = \left[\frac{T_i}{T^e}\left(1+\frac{\lambda_i}{\mu_i}\right)-1\right]\frac{T^e}{T_i}\frac{\mu_i^2}{\lambda_i}.$$

4.3.4. Decomposition equations and DDX algorithm

As said before, we denote by λ_i^e, μ_i^e and T_i^e respectively, the failure rate, the repair rate and the processing time of the machine M_i^e in the equivalent homogenous line. In [51], the authors developed decomposition equations for homogenous lines and propose an efficient algorithm (DDX) to solve these equations.

Production line decomposition methods typically work as follows. An original line is divided into *k-1* lines with only two machines. The method requires the derivation of a set of equations that link the decomposed systems together. Such methods are efficient because systems with two machines can be rapidly analyzed. In general, systems may be represented by discrete or continuous flow models. In both, the processing time is deterministic. The discrete material model has the advantage of better representing the discrete nature of typical factories, but it is limited to systems with equal processing times. The continuous (or fluid) model is better suited in our case because it can be used for systems where the machines have different processing rates. The fluid modelling approach is an approximation which consists in using continuous variables to characterize the flow of parts. Therefore, the quantity of material in each buffer B_i at any time *t* is a real number taking its value in the interval [0, N_i].

The DDX algorithm [51] is the quickest and most reliably convergent algorithm for solving decomposition-type equations. In our optimal design problem, the DDX algorithm can be used to solve the decomposition equations for each configuration. Let recall that in our analytical method the DDX algorithm is applied after approximating each set of parallel machines as a single machine and transforming the resulting non-homogenous production line into an approximate equivalent homogenous line. For more details about DDX algorithm, the reader is referred to [51].

4.4. The hybrid ant colony optimization (HACO) and the simulated annealing

4.4.1. Applying ACS to select machines and buffers: the general algorithm

Following [21], with respect to the problem of selecting machines and buffers in a series-parallel line, each ant is an agent that leaves a pheromone trail, called a trace, on the edges of a graph representing the problem. To represent our problem as such a graph, we introduce the following sets of nodes and edges [56] :

- *Three sets of nodes*:
1. The first set of nodes (N_1) represents the components and the buffers.
2. The second set (N_2) represents the versions of elements available for each component and buffer.
3. The third set of nodes (N_3) represents, for each component, the numbers of elements which can be used in parallel. For example, if the maximum number of elements in parallel is two, the set N_3 will be given by two nodes corresponding to one element and two parallel elements.

- *Two sets of edges*:

1. The first set of edges is used to connect each node in the set N_1 to the corresponding nodes in N_2.
2. The second set of edges is used to connect some nodes in N_2 to the nodes in N_3.

Informally, our algorithm works as follows: m ants are initially positioned on a node representing a component. Each ant represents one possible structure of the entire system. This entire production line structure is defined by the vectors **f**, **h** and **r**. Each ant builds a feasible solution (called a tour) to our problem by repeatedly applying *three* different stochastic greedy rules (i.e., the *state transition rules*). While constructing its solution, an ant also modifies the amount of pheromone on the visited edges by applying the *local updating rule*. Once all ants have terminated their tour, the amount of pheromone on edges is modified again (by applying the *global updating rule*). Ants are guided, in building their tours, by both heuristic information (they prefer to choose "less expensive and more efficient edges"), and by pheromone information.

Note that when an ant builds a solution, it can be feasible or unfeasible. When the obtained solution is unfeasible, it is automatically rejected and it is not taken into account in the comparison with the other feasible solutions obtained by the other ants. It should be noted also that the global update of the pheromone is done only for the best obtained feasible solution.

In the following we discuss the state transition rules, the global updating rule, and the local updating rule.

State transition rules

In the above algorithm, at each step of the construction process, ants use: (1) pheromone trails (denoted by τ_{ij}) to select the versions of machines and buffers and the number of machines connected in parallel; (2) a problem-specific heuristic information (denoted by η_{ij}). The value of η_{ij} depends of the nature of the node (i.e. machine's version or buffer's version). Note that the choice of the number of machines to be connected in parallel is not function of the heuristic information η_{ij}.

An ant positioned on node i (representing a machine or a buffer) chooses the version j ($j= h_i$ if i is a machine and $j = f_i$ if i is a buffer) according to following:

$$j = \begin{cases} arg\ \max_{k \in \Gamma_i} \{(\tau_{ik})^\alpha (\eta_{ik})^\beta\} & if \quad q \leq q_o \\ J & otherwise \end{cases} \qquad (25)$$

where Γ_i is the set of nodes representing the available versions for node i ($\Gamma_i = \{1,...,H_i\}$ if i is a machine or $\Gamma_i = \{1,... ,F_i\}$ if i is a buffer).

And J is a random variable selected according to the probability distribution given by:

$$p_{ij} = \begin{cases} \dfrac{(\tau_{ij})^\alpha (\eta_{ij})^\beta}{\sum\limits_{k \in \Gamma_i} (\tau_{ik})^\alpha (\eta_{ik})^\beta} & if \quad j \in \Gamma_i \\ 0 & otherwise \end{cases} \qquad (26)$$

In the above equations (25) and (26), α and β are parameters that control the relative weight of the pheromone (τ_{ij}) and the local heuristic (η_{ij}), respectively. The value of β depends on

Ant colonies for performance optimization
of multi-components systems subject to random failures
491

the type of a given node. The variable q is a random number uniformly distributed in $[0, 1]$; and q_0 is a parameter $(0 \leq q_0 \leq 1)$ which determines the relative importance of exploitation versus exploration $\eta_{ij} = \dfrac{P(h_i)}{C(h_i)}$ if i represents a machine and $\eta_{ij} = \dfrac{N(f_i)}{C(f_i)}$ represents a buffer.

Similarly, when an ant is positioned on node i representing a version of a machine, it has to select a number j of machines to be connected in parallel. In this case, the used rule is similar to (7) and (8) except for the heuristic information which is set to 1 and $\Gamma_i = \{1,..., R_i\}$.

Global updating rule

Once all ants have built a complete solution, pheromone trails are updated. Only the globally best ant (i.e., the ant which constructed the best design solution from the beginning of the trial) is allowed to deposit pheromone. A quantity of pheromone $\Delta\tau_{ij}$ is deposited on each edge that the best ant has used, where the indices *i* and *j* refer to the edges visited by the best ant. The quantity $\Delta\tau_{ij}$ is given by the total production rate P_{Tbest} of the design feasible solution constructed the best ant. Therefore, the global updating rule is:

$$\tau_{ij} \leftarrow (1-\rho)\tau_{ij} + \rho\Delta\tau_{ij} \tag{27}$$

where $0 < \rho < 1$ is the pheromone decay parameter representing the evaporation of trail.

Global updating is intended to allocate a greater amount of pheromone to greater design solution. Equation (27) dictates that only those edges belonging to the globally best solution will receive reinforcement.

Local updating rule

While building a solution of the problem, ants visit edges on the graph G, and change their pheromone level by applying the following local updating rule:

$$\tau_{ij} \leftarrow (1-\rho)\tau_{ij} + \rho\tau_o \tag{28}$$

where τ_o is the initial value of trail intensities.

The application of the local updating rule, while edges are visited by ants, has the effect of lowering the pheromone on visited edges. This favours the exploration of edges not yet visited, since the visited edges will be chosen with a lower probability by the other ants in the remaining steps of an iteration of the algorithm.

Improving constructed solutions

As said before, it is well known that the performance of ACS algorithms can be greatly improved when coupled to local search algorithms [2]. Following this idea again, an improvement procedure is included in our ACS algorithm, once all ants have terminated their tour and before applying the global updating rule.

This procedure consists of two steps:

1. The remaining budget (the amount not used by the ant) of the obtained structure is first used to improve the solution. In fact, some generated feasible solutions do not use the entire available budget. The procedure improves the initial solution by using this remaining budget to increase the number of machines connected in parallel.

2. In this step, two types of evaluation are performed depending of the nature of the component *(i.e.* machine or buffer). For each pair of components representing the machines, the number of machines is changed by adding one for the first component and subtracting one for the second component. In the case of buffers, the algorithm proceeds to change in turn each pair of chosen versions by another pair.

Thank You for previewing this eBook

You can read the full version of this eBook in different formats:

- ➢ HTML (Free /Available to everyone)

- ➢ PDF / TXT (Available to V.I.P. members. Free Standard members can access up to 5 PDF/TXT eBooks per month each month)

- ➢ Epub & Mobipocket (Exclusive to V.I.P. members)

To download this full book, simply select the format you desire below