

# Real-Time Scheduling in Heterogeneous Systems Considering Cache Reload Time Using Genetic Algorithms

Mohammad Reza Miryani and Mahmoud Naghibzadeh

Department of Computer Engineering, Ferdowsi University of Mashhad,  
Mashhad, Iran  
miryani@stu-mail.um.ac.ir, naghibzadeh@um.ac.ir

**Abstract.** Since optimal assignment of tasks in a multiprocessor system is, in almost all practical cases, an NP-hard problem, in recent years some algorithms based on genetic algorithms have been proposed. Some of these algorithms have considered real-time applications with multiple objectives, total tardiness, completion time, etc. Here, we propose a suboptimal static scheduler of nonpre-emptable tasks in hard real-time heterogeneous multiprocessor systems considering time constraints and cache reload time. The approach makes use of genetic algorithm to minimize total completion time and number of processors used, simultaneously. One important issue which makes this research different from previous ones is cache reload time. The method is implemented and the results are compared against a similar method.

**Keywords:** Hard real-time systems, Heterogeneous multiprocessor systems, Cache reload time, Genetic algorithms, Multi-objective scheduling, Adaptive weight approach.

## 1 Introduction

Computational activities and their responses should be performed within a specified time-frame in real-time systems. A task  $\tau_i$  requested at time  $t_i$  needs  $c_i$  units of time for execution and this time shall be allocated to it before its deadline  $t_i + d_i$ . Otherwise, problems may arise in the system. Real-time systems are classified into two categories with respect to the severity of missing a deadline to Hard Real-Time and Soft Real-Time systems [1]. All deadlines shall meet, in hard real-time systems.

Some optimal schedulers of real-time tasks, on single processor systems to meet deadlines are already developed considering the task characteristics. Scheduling algorithm of EDF and RM are optimal. However, in multiprocessor systems there is no known optimal scheduler [1], [2]. To be optimal for a scheduler means the schedules satisfies one or more criteria of optimality [1], [2]. Generally, it means each task was allocated to a processor such that the overall system is optimal based on predefined criteria. In real-time systems, these criteria can be total tardiness, completion time, throughput, utilization, waiting time, etc. Finding an efficient optimal scheduler for multiprocessor systems is an open problem [3] - [5]. Reference [3] has shown that just

minimizing total tardiness for  $N$  independent task on one machine is an NP-hard problem and in majority of cases, the solution is an NP-hard one. Therefore, developing heuristic algorithms is useable in many applications. Genetic Algorithms is one such algorithm with reasonable efficiency, in many cases [4]-[6]. In recent years, several genetic approaches have been proposed for multiprocessor environments. Reference [6] proposes a scheduler with genetic algorithm for non-preemptive tasks with precedence and deadline constraints but it does not have suitable performance necessarily. Reference [7] presents a hybrid genetic algorithm, in which different operators are applied at different stages of the lifetime, for scheduling partially ordered non-preemptive tasks in a multiprocessor environment. Reference [8] proposes a genetic algorithm implementation to solve a scheduling problem for real-time non-preemptive tasks.

These algorithms minimize only one objective such as completion time, total tardiness, or cost. Reference [5] presents a multiobjective genetic algorithm for scheduling non-preemptive tasks in a soft real-time system with symmetric processors. Nevertheless, some extra local improvement heuristics has been used to find the smallest number of processors. In addition, this work has not considered cache reload time. Other works are in [9], [10] for tasks without timing constraints, but [4] considers to a multiobjective scheduling problem for non-preemptive soft real-time tasks with conflicting objectives, total tardiness and completion time without considering cache reload time. With respect to processor affinity one way to decrease the execution time is to try to assign processors to execute tasks so that two related tasks that share their code and data segments be executed on same processor. Therefore second task does not need to fetch all its data from main memory or auxiliary memories and it can use the already fetched data. In this paper, we propose a new static scheduling algorithm for non-preemptive tasks with precedence constraints on heterogeneous multiprocessor systems, with cache reload time (CRT) and other timing constrains. CRT is important because in practical systems, all timing constrains should be considered otherwise the system may crash. The criteria are completion time and number of processors in a way that all of deadlines are met. Trying to minimize both criteria is done simultaneously. Since there is a conflict between objectives, we use Adaptive Weight Approach [4], [11]. Adaptive Weight Approach uses some useful current population's information in order to justify weights and to move searching in answer space towards positive answers. The rest of the paper is organized as follows. In section 2, scheduling problem for hard real-time tasks on heterogonous multiprocessors will be defined mathematically. Section 3 describes the proposed genetic algorithm, applied procedures, genetic operators, and stopping condition. Section 4 and Section 5 explain validation and the experience results, respectively. Finally, conclusion and future works are in section 6.

## 2 Mathematical Model for Hard Real-Time Scheduling Problem

In this research, we consider the offline scheduling of a set of hard real-time tasks with precedence constraint with task graph on a set of heterogeneous processors in which completion time ( $f_1$ ) and number of processors ( $f_2$ ) are to be minimized under the following conditions:

- All tasks are non-preemptive
- Every task is processed on only one processor at a time

Every processor processes only one task at a given time

- All deadlines must be met.

In addition, there are these assumptions:

- A time unit is an artificial time unit
- Execution time of all tasks on each processor is given
- Precedence relationship or task graph is given prior to scheduling
- Cache reload time with respect to task graph and run time is computable.

Therefore, mathematical statements formulate problem as follows. Presented formulations are developed in [4]-[6] and we have done proper modifications based on new requirements, objectives and limitations of defined problem:

$$\min f_1 = \max \{t_i^F\} \quad (1)$$

$$\min f_2 = \text{Number of Processors} \quad (2)$$

$$s.t. \quad t_i^E \leq t_i^S \quad \forall i \quad (3)$$

$$t_i^E \geq \max_{\tau_j \in \text{pre}(\tau_i)} \{t_j^F\} \quad (4)$$

$$\sum_{m=1}^M x_{im} = 1 \quad \forall i, \quad (5)$$

$$x_{im} \in \{0,1\} \quad \forall i, m \quad x_i = [x_{im}]_{1 \times M} \quad (6)$$

Equations (1) and (2) are fitness functions in this scheduling problem. Equation (1) defines minimizing completion time of tasks because minimization of finish time of each task ( $t_i^F$ ) means that the completion time of the set of tasks is minimized, and (2) expresses minimizing number of processors. Constraint conditions have been shown in (3) to (6). Equation (3) means a task can be started after its own earliest start time begins [4] ( $t_i^S$ : real start time of  $\tau_i$ ). Equation (4) shows earliest start time ( $t_i^E$ ) of the task which is based on its precedence in task graph. In the other word, each task can execute on a processor after its precedence tasks is finished. Equation (5) means that each processor process only one task at a time. Equation (6) is a decision variable because the system is heterogeneous. Note it is required to meet deadlines in hard real-time systems. Thus, there is a default objective formulated as follows:

$$\min f_3 = \sum_{i=1}^N \max \{0, \sum_{m=1}^M (t_i^S + c_{im} - \max \{crt_{ji} x_i x_j^T : \tau_j \in \text{pre}(\tau_i)\} - d_i) x_{im}\} \quad (7)$$

Equation (7) shows that when completion time of task is carried out after the relevant deadline, the system would have tardiness. Otherwise, tardiness will be equal to zero. Tardiness is not acceptable in the hard real-time systems. It is unacceptable because it has deadly effects. So it has to be equal to zero in our study. However, if child task  $\tau_i$  executed on a processor which processed its precedence,  $\max \{crt_{ji} x_i x_j^T : \tau_j \in \text{pre}(\tau_i)\}$  is time will be saved in completion time of  $\tau_i$ .

Following and developing definitions on [4], [5] the following notations are used for the above equations:

- Indices:
  - $i, j$ : task index,  $i, j=1, 2, \dots, N$
  - $m$ : processor index,  $m=1, 2, \dots, M$
- Parameters:
  - $N$ : Total number of tasks
  - $M$ : Total number of processors
  - $G(T, E, K)$ : task graph
  - $T = \{ \tau_1, \tau_2, \dots, \tau_n \}$ : a set of  $N$  tasks
  - $E = \{ e_{ij} \}$ :  $i, j = 1, 2, \dots, N$ , a directed acyclic graph representing precedence relationship
  - $K = \{ k_{ij} \}$ :  $\forall e_{ij} \exists k_{ij}$ : is a random value  $in \in [10^3, 10^6]$
  - $e_{ij}$ : precedence relationship between task  $\tau_i$  and task  $\tau_j$
  - $c_{im}$ : computation time of task  $\tau_i$  on  $m^{\text{th}}$  processor
  - $crt_{ij} = \begin{cases} \gamma k_{ij} \text{ accessTime} & \text{if a same processor processes } \tau_i \text{ and } \tau_j \\ 0 & \text{otherwise} \end{cases}$  (8)
  - $\gamma$ : is a random value in  $[0, 1]$ , *accessTime* is average time to access main memory and auxiliary memory.  $\gamma k_{ij} \text{ accessTime}$  is not more than  $0.05 \times c_{im}$ .
  - $d_i$ : deadline of task  $\tau_i$
  - $pre^*(\tau_i)$ : set of all predecessors of task  $\tau_i$
  - $suc^*(\tau_i)$ : set of all successors of task  $\tau_i$
  - $pre(\tau_i)$ : set of immediate predecessors of task  $\tau_i$
  - $suc(\tau_i)$ : set of immediate successors of task  $\tau_i$
  - $t_i^E$ : earliest start time of task  $\tau_i$
  - $t_i^E = \begin{cases} 0 & \text{if } pre(\tau_i) = \emptyset \\ \max_{\tau_j \in pre^*(\tau_i)} \left\{ t_j^E + \sum_{m=1}^M (c_{jm} x_{jm} - \max \{ crt_{kj} \times x_k x_j^r \mid \tau_k \in pre(\tau_j) \}) \right\} & \text{otherwise} \end{cases}$  (9)
  - $t_i^F$ : finish time of task  $\tau_i$
  - $t_i^F = \min \{ t_i^S + c_{im} x_{im} - \max \{ crt_{\bar{p}} \times x_i x_j^r \mid \tau_j \in pre(\tau_i) \}, d_i \} \forall i$ , (10)
- Decision variables:
  - $t_i^S$ : real start time of task  $\tau_i$
  - $x_{im} = \begin{cases} 1 & \text{if processor } p_m \text{ is selected for task } \tau_i, \\ 0 & \text{otherwise.} \end{cases}$  (11)

### 3 The Proposed Genetic Algorithm

In this paper, our proposed scheduler is based on genetic algorithm. In genetic algorithm, an initial population of feasible answers is shown by a set of chromosomes. Then, a new population of chromosomes is produced by applying operations, such as selection, crossover, mutation, etc. The process of producing new generation continues until a stopping criterion is satisfied. Encoding acts as a mapping of feasible answers space of the problem to initial population and decoding evaluates chromosomes towards an ideal answer. For scheduling problem, several methods and versions for

genetic's operators and procedures have been proposed and some of them can be found in [4]-[6]. In this paper, we propose a new encoding procedure. In addition, we have used the proposed decoding procedure in [4], and have extended it to be useful for our problem.

### 3.1 Encoding

A chromosome  $ch_k = 1, 2, \dots, populationSize$  is a feasible map from set of tasks to set of processors, in which the  $populationSize$  is the total number of chromosomes. A chromosome has two parts:  $u(\cdot)$  and  $v(\cdot)$ .  $u(\cdot)$  shows scheduling order and  $v(\cdot)$  means allocation information [4]. The length of each chromosome is equal to the number of tasks, because all of the tasks must be executed. Scheduling order must satisfy a 'Topologic order' result [12] with respect to task graph. Allocation information determines that each processor shall execute which task. References [4], [5] propose an encoding procedure, while considering topological order but [5] has not implemented topological order. We noticed some errors in the implemented of [4]. In every next level of scheduling, we not only can schedule a task's children but also tasks without precedence. By doing this, we are able to produce more scheduling orders, and it will have positive effect on meeting deadlines. The proposed encoding procedure is shown in Fig. 1. Line 16 is designed to satisfy topological order.

In addition, for initial state, total number of processors is assumed to be equal to the total number of tasks (in worst case). In other hand, in order to meet deadlines, each task must execute on a separate processor, in the worst case. Therefore, in line 13 of Fig.1 M can be equal to N.

```

1  procedure: Encoding for Scheduling Problem for Hard Real-Time
2  input: task graph data set, total number of processors M
3  output:  $u(\cdot), v(\cdot)$ 
4  begin
5       $l \leftarrow 1, W \leftarrow \emptyset;$ 
6      while  $T \neq \emptyset$ 
7           $W \leftarrow \{\tau_i \mid pre^*(\tau_i) = \emptyset \forall i\}$ 
8           $T \leftarrow T - W$ 
9           $j \leftarrow \text{random}(W);$ 
10          $u(l) \leftarrow j;$ 
11          $W \leftarrow W - \{j\}$ 
12          $pre^*(\tau_i) \leftarrow pre^*(\tau_i) - \{\tau_j\} \forall i;$ 
13          $m \leftarrow \text{random}[1: M]$ 
14          $v(l) \leftarrow m;$ 
15          $l \leftarrow l + 1;$ 
16          $T \leftarrow T \cup \{\tau_i, i \in W\};$ 
17     endwhile;
18     output  $u(\cdot), v(\cdot);$ 
19 end;

```

**Fig. 1.** Encoding procedure

### 3.2 Decoding

Decoding procedure is shown in Fig. 2 that is the same decoding in [4]. Total tardiness of each task is computed in line 14, completion time of all tasks is determined in line 18 and number of applied processors is calculated in line 19.

```

1  procedure: Decoding for Scheduling Problem for Hard Real- Time
2  input: task graph data set, chromosome  $ch_k$ 
3  output: schedule set  $S$ , completion time  $f_1$ , number of processors  $f_2$ , total tardiness
    of tasks  $f_3$ 
4  begin
5     $l \leftarrow 1, S \leftarrow \emptyset$ ;
6    while  $l \leq N$ 
7       $i \leftarrow u(l)$ 
8       $m \leftarrow v(l)$ 
9      if (exist suitable idle time) then
10       insert( $i$ );
11     endif;
12     start( $i$ );
13     update_idle();
14      $f_3 \leftarrow f_3 + \max \{ 0, t_i^S + c_{im} - \max \{ crt_{ji} x_i x_j^T : \tau_j \in \text{pre}(\tau_i) \} - d_i \}$ ;
15      $S \leftarrow S \cup \{ i, m : t_i^S - t_i^F \}$ ;
16      $l \leftarrow l + 1$ ;
17   endwhile;
18    $f_1 \leftarrow \max \{ t_i^F \}$ ;
19    $f_2 \leftarrow$  Different Numbers in  $v(\cdot)$ ;
20   output  $S, f_1, f_2, f_3$ ;
21 end;

```

Fig. 2. Decoding procedure

### 3.3 Evolution Function and Competitive Selection

We use Adaptive Weighted Approach (AWA) to dominate conflict between objectives ([4], [11]). In AWA, maximum and minimum values are obtained among all the values of fitness functions of chromosomes by (12). Next, adaptive weight of each fitness function is calculated by (14). Since third objective (minimizing total tardiness) is very essential in hard real-time systems, so we considered priority for each weight. it means for each  $w_p$  ( $p=1,2,3$ ) there is a  $w'_p$  which  $w'_1=0.01$ ,  $w'_2=1$ , and  $w'_3=1000$ . Then, the weighted-sum objective function for each chromosome is computed by (16). Finally, evaluated function for each chromosome is obtained as shown in (17).

For competitive selection we have used of Roulette Wheel Selection [11].

$$f_i^y = \begin{cases} \max\{f_i(ch_k)\} & \text{for } y = \max \\ \min\{f_i(ch_k)\} & \text{for } y = \min \end{cases} \quad (12)$$

where  $i=1,2,3; k=1, \dots, populationSize$  (13)

$$w_p = \frac{1}{f_p^{\max} - f_p^{\min}}, p = 1,2,3 \quad (14)$$

$$w'_1=0.01, w'_2=1, w'_3=1000 \quad (15)$$

$$F(ch_k) = \sum_{p=1}^3 w'_p w_p f_p(ch_k) = \sum_{p=1}^3 (w'_p f_p(ch_k) / (f_p^{\max} - f_p^{\min})) \quad (16)$$

$$eval(ch_k) = 1/F(ch_k) \quad (17)$$

### 3.4 Genetic Operators

We have used modified one-cut crossover and standard mutation ([4], [11]) in this research as shown in Fig. 3.a and 3.b respectively. Procedures of them operate on the  $v(\cdot)$  part of chromosomes. Because, if they operate on the  $u(\cdot)$  part, scheduling order might be changed. Therefore, it will not agree with task graph. So, our modified operators operate only on the  $v(\cdot)$  part of the chromosomes.

<pre> 1  <b>procedure:</b> Crossover 2  <b>input:</b> parent chromosomes <math>ch_k, h_k</math>. 3  <b>output:</b> proto-offspring    chromosomes <math>ch_k, ch_k</math>. 4  <b>begin</b> 5  <math>r \leftarrow random[1:N]</math>; 6  <math>temp \leftarrow v([r+1:N])</math>; 7  <math>v([r+1:N]) \leftarrow v'([r+1:N])</math>; 8  <math>v'([r+1:N]) \leftarrow temp</math>; 9  <b>output</b> <math>ch_k, ch_k</math>; 10 <b>end</b>;</pre>	<pre> 1  <b>procedure:</b> Mutation 2  <b>input:</b> chromosome <math>ch_k</math>, 3  <b>output:</b> offspring chromosomes <math>ch_k</math> 4  <b>begin</b> 5  <math>r \leftarrow random[1:N]</math>; 6  <math>v(r) \leftarrow random[1:M]</math>; 7  <b>output</b> <math>ch_k</math>; 8  <b>end</b>;</pre>
(a)	(b)

Fig. 3. Genetic Operators: (a) Crossover procedure, (b) Mutation procedure

### 3.5 The Proposed Genetic Algorithm

Proposed genetic algorithm has been presented in Fig. 4. Algorithm terminates when main loop in line 7 reaches a default value. In the other hand, it is iterated for a fix number of times.

## 4 Validation

For evaluation of the proposed genetic algorithm several numeral experiments were preformed. Numeral experiments are done with a random precedence task graph.

We used P-Method [13] to produce the random precedence task graph. P-Method is based on an adjacency matrix of a task graph. If there is a precedence relation between tasks  $\tau_i$  and  $\tau_j$  then element  $a_{ij}$  of adjacency matrix will be one, otherwise it will be zero. An adjacency matrix is made with all its lower triangular and diagonal elements equal to zero. Each of the remaining upper triangular elements of the matrix is examined independently as part of a Bernoulli process with factor  $\varepsilon$ , which represents the probability of a success [13], [4], [5]. For the tasks' computation time, deadline and cache reload time between them, we use random numbers based on exponential distribution and normal distribution as same as [4]. However, about value of cache reload time, we assumed it cannot be more than  $0.05 \times \max\{c_{im}^{Exponential} \text{ OR } c_{im}^{Normal}\}$ .

To have suitable compare with reported results in [4] we chose the parameters of genetic algorithm 0.7 for crossover and 0.3 for mutation.

```

1  procedure: Proposed_Genetic_Algorithm
2  input: task graph data set
3  output: best schedule set S
4  begin
5  numberOfGeneration  $\leftarrow$  0;
6  initialize population(numberOfGeneration) byEncoding procedure;
7  while (NumberOfGeneration  $\leq$  maxGeneration) do
8      Evaluate  $f_1, f_2, f_3$  by Decoding procedure;
9      eval(population);           /*eval(chk): k=1,...,populationSize (15)
10     if (not NumberOfGeneration  $\leq$  maxGeneration) then
11         creating new_population by roulette wheel selection;
12         new_population  $\leftarrow$  crossover(new_population);
13         new_population  $\leftarrow$  Mutation(new_population);
14         population  $\leftarrow$  new_population;
15         numberOfGeneration  $\leftarrow$  numberOfGeneration +1;
16     endif;
17 endwhile;
18 output best schedule set S;
19 end;

```

**Fig. 4.** The proposed Genetic Algorithm

## 5 Experiments

Here, we have designed some experiments based on previous sections. We compared our proposed algorithm with the proposed algorithm in [4] because it acceded better answers in comparison the others.

### 5.1 Experiment 1

The first experiment is taken from [4]. In this experiment we have some information as well as it is shown in Table 1 that has been created by the P-Method.



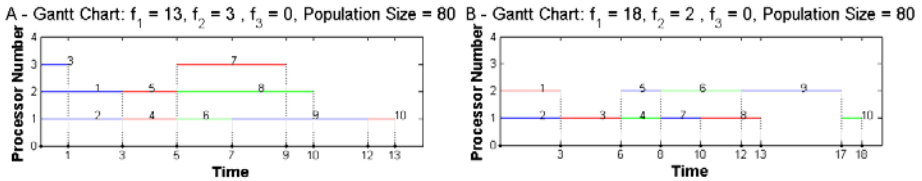
**Table 1.** Data set of experiment 1

$i$	$pre^*(\tau_i)$	$c_{im}$			$d_i$
		$c_{i1}$	$c_{i2}$	$c_{i3}$	
1	8	5	3	10	13
2	6	3	7	12	17
3	4, 5	3	4	1	12
4	6, 7, 8	2	16	6	12
5	6, 10	12	2	4	27
6	9	2	4	7	24
7	-	2	15	4	13
8	-	3	5	4	18
9	10	5	5	8	27
10	-	1	5	6	29

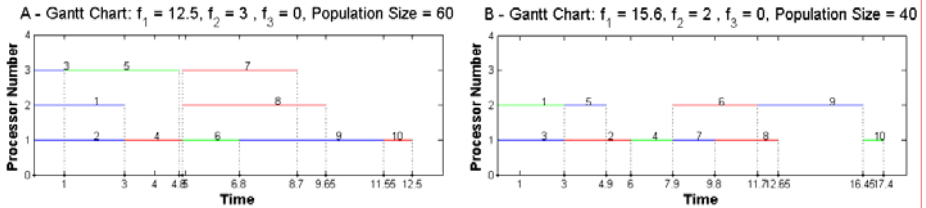
**Table 2.** Data set of experiment 2

$i$	$pre^*(\tau_i)$	$c_{im}$				$d_i$
		$c_{i1}$	$c_{i2}$	$c_{i3}$	$c_{i4}$	
1	8	5	3	11	8	19
2	6	6	5	4	13	9
3	4, 5	11	8	6	7	18
4	6, 7, 8	10	13	5	6	37
5	6, 10	10	13	8	11	38
6	9	2	11	11	3	37
7	-	3	10	11	8	44
8	-	4	12	10	5	30
9	10	6	9	7	10	37
10	-	11	12	6	4	58

We divided this experiment into two parts. In first part we scheduled it without considering cache reload time. In our method, the best answer as the completion time ( $f_1$ ) is 13, total tardiness ( $f_3$ ) is 0, and number of applied processors ( $f_2$ ) is 3 where asreport of this experiment in [4] is as completion time; 15, total tardiness is 6 and minimizing of number of processors is not their objective (Fig. 5-A). In addition, our proposed algorithm could schedule this task graph with 2 processors as is shown in Fig. 5-B. In the second part with respect to cache reload time, two suboptimal schedulers are shown in Fig. 6. In addition, some obtained results are shown in Fig. 7 and Fig. 8. In Fig. 7 has been shown a comparison between our proposed method and the proposed method in [4]. In Fig. 8 we computed average of completion time of best obtained scheduler in 50 iterations of experience with considering and nonconsidering cache reload time when population size increases.



**Fig. 5.** Experiment 1 without cache reload time



**Fig. 6.** Experiment 1 with cache reload time

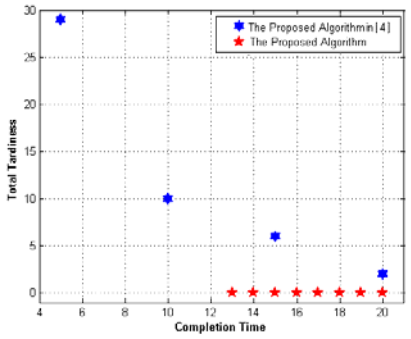


Fig. 7. A comparison between our proposed method and the proposed method in [4]

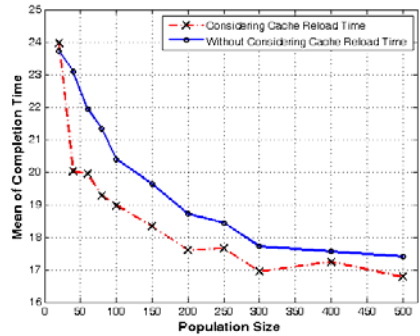


Fig. 8. A comparison of completion time considering and nonconsidering cache reload time

### 5.2 Experiment 2

Second experiment was taken from [4] too. Execution time of each task on processors is given in Table 2. Fig. 9 shows three instances of the best output answers that they describe the best answer of our proposed method according to Table 2 without considering cache reload time. In Fig. 10 has been shown two schedulers considering cache reload time. It is obvious scheduler with 3 processors has a lower completion time than scheduler with 4 processors. In fact this Comparison shows importance of considering cache reload time. An overall result has been presented in Fig. 11.

Table 3. Data set of experiment 2 with duplicated processors

<i>i</i>	$pre^*(\tau_i)$	$C_{im}$								$d_i$
		$C_{i1}$	$C_{i2}$	$C_{i3}$	$C_{i4}$	$C_{i5}$	$C_{i6}$	$C_{i7}$	$C_{i8}$	
1	8	5	3	11	8	5	3	11	8	19
2	6	6	5	4	13	6	5	4	13	9
3	4, 5	11	8	6	7	11	8	6	7	18
4	6, 7, 8	10	13	5	6	10	13	5	6	37
5	6, 10	10	13	8	11	10	13	8	11	38
6	9	2	11	11	3	2	11	11	3	37
7	-	3	10	11	8	3	10	11	8	44
8	-	4	12	10	5	4	12	10	5	30
9	10	6	9	7	10	6	9	7	10	37
10	-	11	12	6	4	11	12	6	4	58

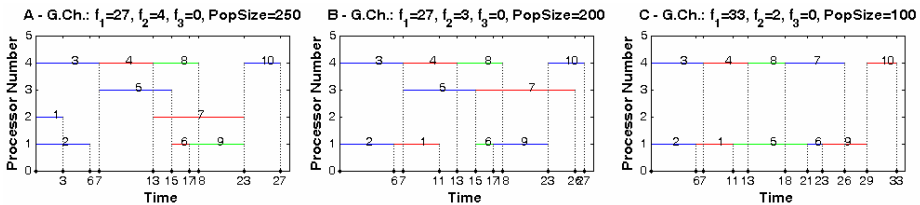


Fig. 9. Schedulers for experiment 2 considering Table 5 and without considering cache reload time

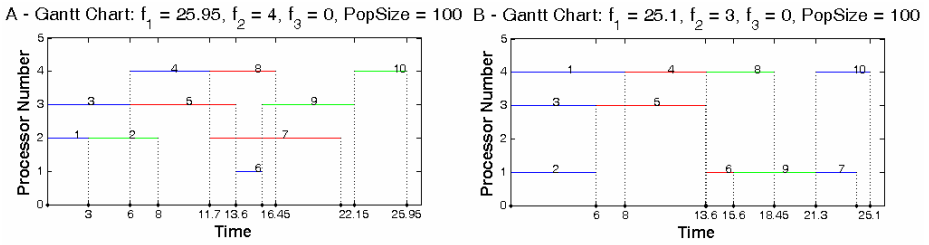


Fig. 10. Two best schedulers considering cache reload time

### 5.3 Experiment 3

In this part for testing of processors redundancy of proposed algorithm, we duplicated the assumed processors from Table 2 to Table 3 Overall results has been shown in Fig. 11 which this figure declares not only considering cache reload time in scheduling can cause using lower processor and decreases completion time simultaneously but also our proposed algorithm could gain obtained results when we use data set Table 2 without duplication. Fig. 12 shows always increasing number of processors does not lead to better answer for completion time where there is not tardiness.

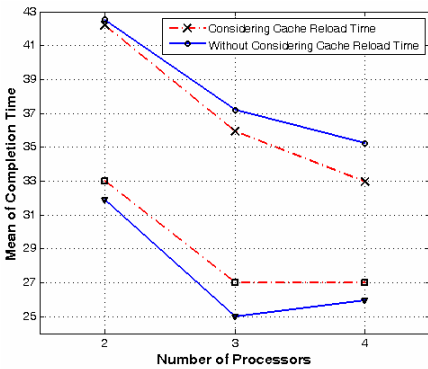


Fig. 11. Comparisons between obtained completion time considering and nonconsidering cache reload time

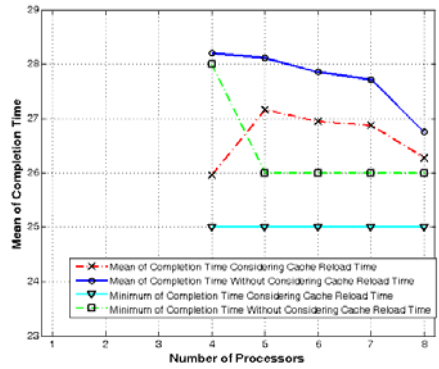


Fig. 12. Comparisons between obtained completion time considering and nonconsidering cache reload time

## 6 Conclusion and Future Works

In this paper we have tried generalization on latest works ([4] – [6]) and covered their shortcomings. Paying attention to cache reload time in heterogeneous real-time systems is one of the aspects of this work.

Improving the encoding procedure which has large influence on the proposed scheduler is part of this paper. For this part, we designed a suitable encoding procedure based on topological sort. Same chromosome like [4] has been used to be able for comparing the proposed algorithm and the proposed algorithms in [4].

Also, trying to minimize the number of heterogeneous processors while all deadlines are met is done by genetic algorithm, whereas, in [5] this work is done by some extra local improvement heuristics moreover we have considered heterogeneous processors as oppose to [5] which assumes homogeneous processors. So our proposed algorithm is a generalization on [5]. Unlike [6], designed chromosome is simple and efficient, has fewer limitations, and while using limited information can minimize conflicted objective simultaneously.

For future works, we will try to design some better stopping conditions, perform some improvements on convergent conditions, and compare with PSO method.

## References

1. Krishna, C.M., Kang, G.S.: Real-time systems. McGraw-Hill, New York (1997)
2. Marwedel, P.: Embedded System Design. Springer, Netherland (2006)
3. Du, J., Leung, J.Y.T.: Minimizing total tardiness on one machine is NP-hard. *Mathematics of Operational Research* 15, 483–495 (1990)
4. Yoo, M., Gen, M.: Scheduling algorithm for real-time tasks using multiobjective hybrid genetic algorithm in heterogeneous multiprocessors system. *Journal of Computers & Operations Research* 34, 3084–3098 (2007)
5. Oh, J., Wu, C.: Genetic-algorithm-based real-time task scheduling with multiple goals. *Journal of Systems and Software*, 245–258 (2004)
6. Mitra, H., Ramanathan, P.A.: Genetic approach for scheduling non-preemptive tasks with precedence and deadline constraints. In: 26th Hawaii international conference on system sciences, pp. 556–564 (1993)
7. Lin, M., Yang, L.: Hybrid genetic algorithms for scheduling partially ordered tasks in a multi-processor environment. In: Sixth international conference on real-time computer systems and applications, pp. 382–387 (1999)
8. Monnier, Y., Beauvais, J.P., Deplanche, A.M.: A genetic algorithm for scheduling tasks in a real-time distributed system. In: 24th euromicro conference, pp. 708–714 (1998)
9. Page, A.J., Naughton, T.J.: Dynamic task scheduling using genetic algorithm for heterogeneous distributed computing. In: 19th IEEE international parallel and distributed processing symposium, p. 189.1 (2005)
10. Dhodhi, M.K., Ahmad, I., Ahmad, I., Yatama, A.: An integrated technique for task matching and scheduling onto distributed heterogeneous computing systems. *Journal of Parallel and Distributed Computing* 62, 1338–1361 (2002)
11. Gen, M., Cheng, R.: Genetic Algorithm and Optimization Engineering. John Wiley and Sons, INC., New York (2000)
12. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 2nd edn. McGraw-Hill, New York (2001)
13. Al-Sharaeh, S., Wells, B.E.: A comparison of heuristics for list schedules using the box-method and P-method for random digraph generation. In: 28th Southeastern symposium on system theory, pp. 467–471 (1996)
14. Cosnard, M., Marrakchi, M., Robert, Y., Trystram, D.: Parallel Gaussian elimination on an MIMD computer. *Journal of Parallel Computing* 6(3), 275–296 (1998)
15. Wu, M.Y., Gajski, D.D.: Hypertool: A programming aid for message-passing system. *IEEE Trans. Parallel and Distributed Systems* 1(3), 330–343 (1990)