



Mixed-criticality real-time scheduling of gang task systems

Ashikahmed Bhuiyan¹ · Kecheng Yang² · Samsil Arefin³ ·
Abusayeed Saifullah⁴ · Nan Guan⁵ · Zhishan Guo¹

Accepted: 24 April 2021 / Published online: 23 May 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Mixed-criticality (MC) scheduling of sequential tasks (with no intra-task parallelism) has been well-explored by the real-time systems community. However, till date, there has been little progress on MC scheduling of parallel tasks. MC scheduling of parallel tasks is highly challenging due to the requirement of various assurances under different criticality levels. In this work, we address the MC scheduling of parallel tasks of gang model that allows workloads to execute on multiple cores simultaneously, as well as the change to degree of parallelism of a task upon a mode switch. It represents an efficient mode-based parallel processing scheme with many potential applications. To schedule such task sets, we propose a new technique GEDF-VD, which integrates Global Earliest Deadline First (GEDF) and Earliest Deadline First with Virtual Deadline (EDF-VD). We prove the correctness of GEDF-VD and provide a detailed quantitative evaluation in terms of speedup bound in both the MC and the non-MC cases. Specifically, we show that GEDF provides a speedup bound of 2 for non-MC gang tasks, while the speedup for GEDF-VD considering MC gang tasks is $\sqrt{5} + 1$. Experiments on randomly generated gang task sets are conducted to validate our theoretical findings and to demonstrate the effectiveness of the proposed approach.

Keywords Multi-core systems · Mixed criticality · Real-time scheduling · Speedup bound

1 Introduction

Due to size, weight, and power considerations, there is a trend that multiple tasks with different criticality levels (that are subject to varying degrees of assurance/verification) share a computing platform (Vestal 2007). This type of system is commonly

✉ Zhishan Guo
zsguo@ucf.edu

Extended author information available on the last page of the article

known as a *mixed-criticality (MC) system*, where each task can be associated with various execution budgets. During normal operation, all tasks are scheduled according to their typical execution budget. However, some critical tasks may exceed their typical budget and need more resources to finish their execution. Suppose the available resources are not sufficient in these scenarios. In that case, the less critical task will be sacrificed to free up the resources for accommodating the additional computational requirements requested by the more critical ones.

Take an avionics software standard as an example, where the ground control subsystems are more safety-critical than ground communication and light controls. During the incident of emergency (e.g., an accident), it is more important to execute the safety-critical components rather than the other components. On the other hand, in normal condition, all these components are required to perform smoothly (for more details, refer to the Table 1.1 of Li (2013), which demonstrates the RTCA DO-178B avionics software standard). MC scheduling has received considerable attention (refer to Burns and Davis (2013) for a thorough and updated survey) as it brings significant improvements in resource efficiency.

Note that safety-critical MC systems have tight correctness requirements. These requirements can be verified by two related but orthogonal perspectives: *a priori verification* and *run-time robustness* (Baruah 2018). Before run-time, a priori verification determines whether a system will behave correctly (or not) during execution, while run-time robustness deals with unexpected system behavior at run-time. There are some debates regarding MC's applicability to run-time robustness (Ernst and Di Natale 2016; Esper et al. 2015). Although an MC system has imitation from the perspective of a priori verification (which is our work scope), these criticisms usually do not hinder the applicability of an MC system for its designed scope of a priori verification (Baruah 2018).

Parallel computing workloads Recent advances in parallel computing allow executing a single piece of code simultaneously on multiple computing units or a set of threads to execute on multiple processors concurrently. Such design provides a much better capability of exploiting the benefits provided by modern platforms. As a result, there is an urgent need in handling workload models that allow intra-task parallelism (i.e., parallel tasks). Parallel computing systems perform a large number of computations and often need to interact with their surroundings under real-time constraints, e.g., arms system (RADAR). In these types of applications, a lot of processors co-operate with each other, and these communications are timing critical. It is necessary for a system to have both high performance and predictability; i.e., efficient control that minimizes the introduced overhead, while responding to external events (coming through sensors) in real-time. In this work, we consider the gang scheduling, where all threads of a task are grouped into a gang, and during execution, the whole group is concurrently scheduled on distinct cores. The gang task model is a practical, widely used, and representative workload model for intra-task parallelism (Alahmad and Gopalakrishnan 2019; Dong and Liu 2017; Goossens and Richard 2016; Kato and Ishikawa 2009). Also, the gang task model is supported by some widely used parallel computing programming standards [e.g., openacc (2017)], which is commonly used in graphics processing unit (GPU).

Existing work The real-time systems and parallel computing communities have given considerable attention towards these two directions: *MC scheduling* and *scheduling of parallel tasks*. These two emerging trends bring in some critical and exciting problems, and there is an emerging need in integrating those two trends. There has been extensive research on the (a) MC scheduling of sequential (i.e., non-parallel) tasks (refer to the recent survey in Burns and Davis 2018, 2013) and (b) scheduling of parallel tasks with a single-criticality level (Kato and Ishikawa 2009; Goossens and Richard 2016; Dong and Liu 2017; Andersson and de Niz 2012; Bonifaci et al. 2013; Li et al. 2014). Till date, very few efforts (Liu et al. 2014; Baruah 2016; Li et al. 2017) have been made towards the combined problem of MC scheduling of parallel tasks. To our knowledge, none of these efforts has considered mixed-criticality gang task scheduling on multi-core platforms.

Motivation behind this work Multi-core platform enables applications that require better energy efficiency, higher performance, and real-time guarantees. The notion of MC systems with the intra-task parallelism stems from many current trends. For example, the number of cores fabricated on a chip is increasing rapidly. Besides, the computational demand for an individual task (with stringent timing requirements) is rising, which makes it essential to consider the intra-task parallelism. Furthermore, when safety-critical and non-safety-critical tasks share a common computational platform, there is an increasing demand to integrate functionality with different criticality levels. Such demand promotes the idea of MC scheduling, i.e., combining various functionalities of varying criticality levels onto the same computing platform.

In this work, we study the mixed-criticality gang task scheduling, and the fault-tolerant systems would be a promising application for such study. A fault-tolerant system often follows the mixed-criticality model (Burns and Davis 2013). If a fault is identified in such a system, it is recovered via various recovery techniques, e.g., exception handling, recovery blocks, and task replication. Some extra work has to be undertaken upon identifying a fault, which leads to the abandonment or delay of some less critical works. The impact of a fault in such a system ranges from no visible effect to an entire system crash. To overcome these faults, the *ASTEROID* project is proposed (a cross-layer fault-tolerance solution for the mixed-criticality platform) that detects errors and recovers the system in different software and hardware layers (Döbel et al. 2012; Engel and Döbel 2012). Considering such a cross-layer platform, a recent work in Rambo and Ernst (2017) proposed a replica-aware co-scheduling (with Strict Priority Preemptive protocols) for a mixed-criticality system that improves the system performance. They have considered the replicas as a gang that is activated concurrently on multiple cores.

Challenges In gang task model, a task cannot start execution until the number of available cores is no less than what is required by it (i.e., a task's *degree of parallelism*). This simple constraint adds a huge restriction on real-time schedulability and makes the problem highly challenging. We are aware of only one known correct schedulability analysis (Dong and Liu 2017) under Global Earliest Deadline First (GEDF) for gang tasks. Besides, integrating MC in gang scheduling scheme adds additional challenges due to the dual notion of correctness. In the normal mode, a task may have a utilization less than 1, while in the critical mode, the utilization

could be much higher than one Dong and Liu (2017). Such a change in the utilization adds significant complexity in speedup bound analysis (Baruah et al. 2015a). For example, in the speedup bound analysis for MC scheduling of ordinary sporadic tasks, an individual task's utilization is at most the processor's speed is a straight forward and necessary feasibility condition. At the same time, it no longer holds for the gang tasks. Besides, the scheduler does not know the exact behavior of each task before run-time (non-clairvoyant). Hence, the scheduler must be able to detect the critical condition early enough to allocate more resources to the more critical tasks to handle this drastic change and still be able to meet the deadlines.

This research In this paper, we study the real-time scheduling of MC gang tasks on identical multi-core platforms. We propose the first scheduling algorithm GEDF-VD (GEDF with Virtual Deadline) for MC gang tasks. Our approach leverages the synthesis of uniprocessor scheduling techniques such as EDF-VD (Baruah et al. 2012) as well as *GEDF* (Dong and Liu 2017) that was designed for non-MC gang tasks. To our knowledge, this is the first work that studies the MC scheduling of the gang task model. Specifically, we make the following contributions:

- We generalize the gang task model to the MC context by incorporating extensions on both the execution time and the degree of parallelism dimensions, and propose GEDF-VD for the generalized model. We also conduct a utilization based schedulability test and prove its correctness formally.
- We prove that the speedup bound (Baruah et al. 2015a) for GEDF to gang tasks in a non-MC platform is at most $(2 - 1/(M + 1 - \min_i\{m_i\}))$, where M denotes the total number of processor cores and m_i denotes the degree of parallelism of task τ_i . To our knowledge, this is the first speedup bound result for GEDF scheduling of gang tasks.
- With the result from the previous step, we then derive a speedup bound of $\sqrt{5} + 1$ for GEDF-VD considering MC gang tasks.
- Extensive simulations under randomly generated task sets are conducted to demonstrate the real-time performance and effectiveness of the proposed algorithm in terms of acceptance ratio, which is defined as the ratio of the number of schedulable task sets over the total number of task sets.

Organization The remainder of this paper is organized as follows. Section 2 describes the task model, notations, and preliminaries. Section 3 provides a detailed description of our scheduling algorithm and prove its correctness. Section 4 derives the speedup bounds for the non-MC and MC platform, under GEDF and GEDF-VD scheduling algorithms, respectively. Simulation results are presented in Sect. 5. Section 6 discusses related prior work. Section 7 concludes this paper and points out future research directions.

2 Dual-criticality gang task model

In this work, we consider the problem of scheduling a task set $\tau = \{\tau_1, \dots, \tau_n\}$ of n independent *implicit deadline* (i.e., the period of a task is equal to its deadline) sporadic MC gang tasks on M identical cores. In this model, each task generates an infinite number of MC gang jobs (the j th job of task τ_i is denoted as $\tau_{i,j}$). To describe the dual-criticality gang task model, first, we provide details on traditional *non-MC gang task model* and *MC sporadic sequential task model*. Then, by leveraging these two models, we generalize the gang task model to the MC context. In this work, we restrict our attention to dual-criticality model, and consider it is an important step towards the multi-criticality systems.

Non-MC gang task model In traditional non-MC gang task model, each task τ_i is represented with a 4-tuple (m_i, c_i, T_i, D_i) , where m_i is the degree of parallelism and each job of task τ_i requires access to m_i cores for at most c_i time units to complete its execution, T_i is the task period, and D_i is the relative deadline. In gang scheduling, each task consists of multiple threads (referred to as a gang), and each thread of the same task occupies a processor for the same time quantum (Kato and Ishikawa 2009; Dong and Liu 2017). Hence, in the time-space, the execution requirement of any job $\tau_{i,j} \in \tau_i$ can be represented as an $m_i \times c_i$ rectangle. The relative deadline D_i specifies that for each of the released jobs $\tau_{i,j}$ (of task τ_i), its deadline $d_{i,j} = r_{i,j} + D_i$, where $r_{i,j}$ denotes the release time of $\tau_{i,j}$ (Dong and Liu 2017). The *utilization* u_i of each task $\tau_i \in \tau$ is given by $u_i = (m_i c_i) / T_i$, and the overall system utilization is: $U_{sum} = \sum_{\tau_i \in \tau} u_i$. Note that, it is possible that the value of u_i is larger than one, which is different from the traditional sequential task model. Based on the scheduling flexibility, a gang task τ_i can be categorized into three groups. A task τ_i is said to be:

- *rigid*, if m_i is fixed a priori and does not change throughout the execution,
- *moldable*, if m_i is fixed during its activation and does not change throughout the execution,
- *malleable*, if m_i is not fixed and can be changed during its execution by the scheduler.

We currently restrict our attention to the *rigid* task model. This model suits various applications that use parallelism, some of which are implemented using the message-passing approach and tools like MPI.

MC sporadic task model In a dual-criticality systems, the criticality level of τ_i is represented by $\chi_i = \{lo, hi\}$. The worst case execution time (WCET) estimations of each task is also represented by a tuple (c_i^{lo}, c_i^{hi}) where c_i^{lo} and c_i^{hi} represent the LO and HI-criticality WCETs respectively. c_i^{hi} is measured by a more pessimistic tool by considering all possible scenarios, while c_i^{lo} is calculated using a less pessimistic yet realistic tool. Collection of all LO- and HI-criticality tasks in τ are denoted by τ_{lo} and τ_{hi} respectively. u_i^{lo} and u_i^{hi} denotes the utilization of τ_i in LO- and HI-criticality mode respectively, where $u_i^{lo} = c_i^{lo} / T_i$ and $u_i^{hi} = c_i^{hi} / T_i$.

MC gang task model By leveraging the above two models, in our work, we consider a workload model of MC gang tasks, where each task τ_i is represented by a 7-tuple $(m_i^{lo}, m_i^{hi}, \chi_i, c_i^{lo}, c_i^{hi}, T_i, D_i)$, where

- $m_i^{lo} (m_i^{hi})$ = degree of parallelism of task τ_i in LO(HI) criticality mode, and $m_i^{lo} \leq m_i^{hi}$.
- χ_i = criticality level of each task τ_i and $\chi_i \in \{ lo, hi \}$
- $c_i^{lo} (c_i^{hi})$ = τ_i 's WCET in LO(HI)-criticality mode, and $c_i^{lo} \leq c_i^{hi}$.
- T_i = minimum inter-arrival time between jobs.
- D_i = relative deadline.

In this model, a task τ_i occupies $m_i^{lo} (m_i^{hi})$ processors for $c_i^{lo} (c_i^{hi})$ time quantum at LO(HI)-criticality mode. Note that, if $\forall \tau_i, m_i^{lo} = m_i^{hi} = 1$, i.e., degree of parallelism for each gang task is 1, our analysis (Sects. 3 and 4) will reduce to the existing MC scheduling method designed for the sporadic task model. We believe this is common for a restricted special case of a more complex and expressive model. For example, the directed acyclic graph (DAG) task model (Bonifaci et al. 2013; Li et al. 2014) is popular to represent intra-task parallelism. Many of the existing schedulability analysis considering the DAG model would also reduce to prior study for ordinary sporadic tasks if the number of nodes of each DAG task is equal to 1.

Now, we generalize the utilization concepts to suit the MC gang task model, which are analogous to the above-mentioned concepts. Refer to the Example 1 for details.

$$\begin{aligned}
 U_{lo} &\stackrel{\text{def}}{=} \sum_{\tau_i \in \tau_{lo}} m_i^{lo} \times c_i^{lo} / T_i, \\
 U_{hi} &\stackrel{\text{def}}{=} \sum_{\tau_i \in \tau_{hi}} m_i^{lo} \times c_i^{lo} / T_i, \\
 U_{hi} &\stackrel{\text{def}}{=} \sum_{\tau_i \in \tau_{hi}} m_i^{hi} \times c_i^{hi} / T_i
 \end{aligned}$$

Example 1 Consider the task-set $\tau = (\tau_1, \tau_2, \tau_3)$ in Table 1. For this task-set we derive the utilization as follows:

$$U_{lo} = c_2^{lo} \times m_2^{lo} / T_2 = 0.6, U_{hi}^{lo} = c_1^{lo} \times m_1^{lo} / T_1 + c_3^{lo} \times m_3^{lo} / T_3 = 2, \text{ and } U_{hi}^{hi} = c_1^{hi} \times m_1^{hi} / T_1 + c_3^{hi} \times m_3^{hi} / T_3 = 3.6.$$

Example 2 Consider the MC gang task set in Table 1 to be scheduled in four cores. A GEDF schedule for this task set is shown in Fig. 1. The system starts at LO-criticality mode, and all the tasks (τ_1, τ_2, τ_3) will execute up-to C_i^{lo} . Recall that, m_i^{lo} is the degree of parallelism of τ_i in LO-criticality mode. Hence, τ_1 cannot execute at $t = 5$ as it needs three cores to execute while only two cores (P_3 and P_4) are idle. At a mode switch ($t = 9$), all LO-criticality tasks (τ_2) are dropped, and all HI-criticality tasks (τ_1, τ_3) will execute up-to C_i^{hi} (at m_i^{hi} cores). After a mode switch, all

Table 1 An MC gang task set with GEDF schedule shown in Fig. 1

Task ID	c_i^{lo}	c_i^{hi}	T_i	χ_i	m_i^{lo}	m_i^{hi}
τ_1	3	4	5	HI	3	4
τ_2	3	3	10	LO	2	2
τ_3	1	2	10	HI	2	2

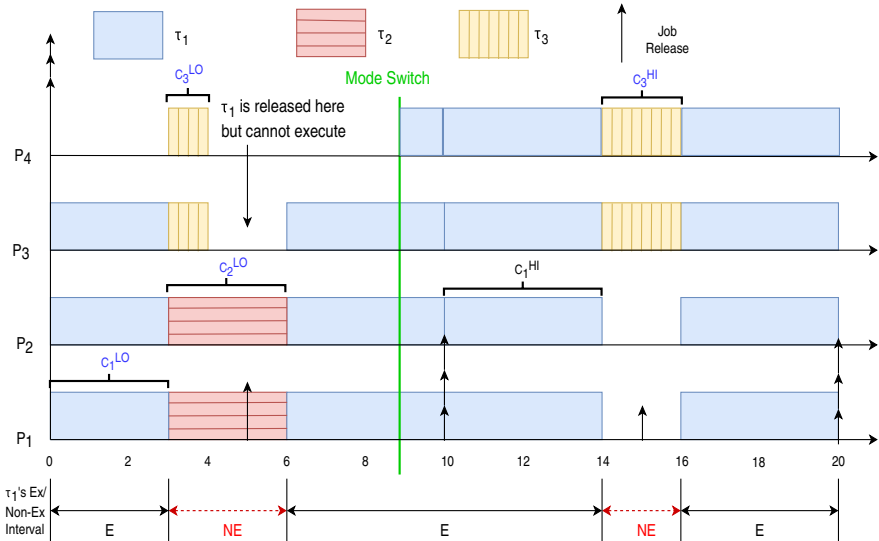


Fig. 1 A GEDF scheduling of the MC gang task-set from Table 1 and the executing (E)/non-executing (NE) intervals of τ_1

HI-criticality jobs (including the ones which are currently executing) will execute up-to their HI-criticality WCET.

Motivations behind this model Some commonly used parallel computing programming standards [e.g., openacc (2017)] support the gang task model. OpenACC is one of the parallel computing programming standards used for GPU architecture. Recently, there has been extensive research on GPU architecture (few to mention Capodieci et al. 2018; Elliott et al. 2013; Xiao and Feng 2010; Yang et al. 2018). GPU architecture is popular because of the features like (1) highly threaded but low context switch latency architecture, (2) high parallelism and (3) minimal dependency between data elements, etc. Previous work on GPU scheduling considered limited or no preemption policy (Elliott et al. 2013; Xiao and Feng 2010). However, this work is motivated by some recent attempts to incorporate the *preemptive* support in GPUs. For example, a prototype has been implemented and tested with preemptive support (at the pixel level and the thread level) in a virtualized environment in a recent work (Capodieci et al. 2018). Its prototype is EDF based, and enhanced

with a bandwidth isolation mechanism (e.g., constant/total bandwidth servers (Spuri and Buttazzo 1994) for the graphics and computing workloads. Also, the prototype is tested on a recent NVIDIA Tegra-based system on a chip (SoCs) (nvidia 2017). Since some recent work study the preemptive support in the GPU architecture, there is a need for a comprehensive study of gang task scheduling using GEDF.

Now, we introduce some definitions and preliminaries which will be frequently used in later sections of this paper.

Definition 1 (*MC-correct schedule*) Scheduling strategy must ensure an *MC-correct* schedule, as defined below (Li et al. 2017).

- If the system stays in normal condition (i.e., each task in the system finishes execution within its LO-criticality WCET), all tasks must meet their deadlines.
- If the system transits into a critical condition (i.e., there exists a HI-criticality task executing beyond its LO-criticality WCET), all HI-criticality tasks must meet their deadlines, while LO-criticality tasks need not so.

Definition 2 (*Executing/Non-Executing interval*) An interval $[t_1, t_2)$ (where $t_1 < t_2$) is an **executing** interval for a task τ_i if m_i out of M cores are executing the current active job released by τ_i throughout this interval. Otherwise, $[t_1, t_2)$ is a **non-executing** interval for τ_i . An illustrative example is shown in Fig. 1 by pointing the executing and non-executing intervals for task τ_1 .

Definition 3 (*Active/pending task*) If there exists a task $\tau_i \in \tau$, such that it has a job τ_{ij} where $r_{ij} \leq t < d_{ij}$. Here, r_{ij} and d_{ij} respectively denotes the release time and deadline of τ_{ij} , then τ_i is considered as an active task at time t . A job is *pending* if it is released but not finished (Dong and Liu 2017).

Definition 4 Maximum possible number of idle cores Δ_i^{lo} (Δ_i^{hi}) for a task τ_i refers to the maximum number of available cores (that are not executing any job) at any time in LO (HI-criticality mode) during τ_i 's non-executing intervals in which it has a pending job (Dong and Liu 2017).

Example 3 Let us consider a 4-core platform and task set $\tau = \{\tau_1, \tau_2, \tau_3\}$ from Table 1. At LO-criticality mode, the degree of parallelism for these tasks are given as: $m_1^{\text{lo}} = 3$, $m_2^{\text{lo}} = m_3^{\text{lo}} = 2$. For this task set, the maximum possible number of idle cores at LO-criticality mode is: $\Delta_1^{\text{lo}} = 2$, $\Delta_2^{\text{lo}} = \Delta_3^{\text{lo}} = 1$. This is because τ_1 cannot execute at time t (even when it has a pending job) if τ_2 or τ_3 is executing at t . The degree of parallelism for τ_2 (or τ_3) is 2. So, the maximum number of idle cores for τ_1 is Δ_1 , where $\Delta_1 = 4 - 2 = 2$. We can calculate the value of Δ_2 and Δ_3 in the same approach (refer to Algorithm 1 in Dong and Liu (2017)). At HI-criticality mode, degree of parallelism for these tasks becomes: $m_1^{\text{hi}} = 4$, $m_2^{\text{hi}} = m_3^{\text{hi}} = 2$, and the maximum possible number of idle cores will be: $\Delta_1^{\text{hi}} = 2$, $\Delta_2^{\text{hi}} = \Delta_3^{\text{hi}} = 0$.

System behavior and scope of this work It is expected that an MC system starts execution in normal mode. The system-wide mode transition is triggered if a HI-criticality task τ_i has received cumulative execution length beyond its LO-criticality WCET and did not signal its finishing. Like the Vestal model (Vestal 2007), after a mode switch, no LO-criticality tasks get any service guarantee. After mode transition (from LO-criticality to HI-criticality), at the first idle instant, the system switches back to the LO-criticality mode again. All other scenarios (e.g., a HI-criticality task runs for more than its HI-criticality WCET) are considered as *erroneous*, where no guarantees will be made and hence is not considered.

3 GEDF-VD for dual-criticality system

Now we describe our algorithm for the MC task systems considering the GEDF-VD algorithm. In this work, we consider implicit deadline (So, we use the terms *deadline* and *period* interchangeably) sporadic task systems on identical multi-core platforms. We integrate an uniprocessor MC scheduling technique [EDF-VD (Baruah et al. 2012)] with a multiprocessor gang task scheduling technique [GEDF (Dong and Liu 2017)] and derive a new algorithm named GEDF-VD (Sects. 3.1 and 3.2). In our approach, we determine a *scaling factor*, which scales the deadline of all HI-criticality tasks at LO-criticality mode. This factor will be calculated in such a way that the correctness of the system can be guaranteed at both LO- and HI-criticality modes (Sects. 3.3 and 3.4).

3.1 EDF-VD and GEDF-VD: an overview

EDF-VD In case of a mode switch (LO to HI), to generate an *MC-correct schedule* (Definition 1), a scheduler must ensure that all HI-criticality tasks meet their deadlines (while LO-criticality tasks can be sacrificed). To guarantee this criterion, a specific amount of CPU time must be reserved for those HI-criticality tasks even if the system is running at LO-criticality mode. This reservation of time can be achieved by shortening the deadlines of HI-criticality tasks under normal mode—those are virtual deadlines.

In EDF-VD, deadlines of all HI-criticality tasks are shortened by multiplying them with a *scaling factor*, and this updated deadline is called the *virtual deadline*. During run-time (at LO-criticality mode), all HI-criticality (LO-criticality) tasks are executed according to their virtual (actual) deadline, according to the EDF order. Upon a mode switch, only the HI-criticality tasks are executed in the EDF order w.r.t their actual/original deadlines.

In the case of a LO- to HI-criticality mode-switch, a HI-criticality task demands additional computational requirements. Setting a virtual deadline for the HI-criticality tasks leaves enough time to finish the extra workload within their actual deadlines. If the virtual deadline is too short, it increases the system density at

normal (i.e., LO-criticality) mode. In contrast, a large virtual deadline threatens the schedulability of the system after a LO- to HI-criticality mode switch. The trick is to determine a balanced scaling factor x , such that the correctness under both execution modes can be guaranteed. Baruah et al. (2012) showed the steps to calculate the minimum x that guarantees the schedulability of all tasks in the system. They also proved the improvement in system schedulability by reducing the deadline for HI-criticality tasks at LO-criticality mode.

Remark 1 In this work, we consider a completely different Gang task workload model in a multi-core platform. As a result, the approach proposed by Baruah et al. (2012) to calculate the scaling factor x , as well as the schedulability test, are no longer applicable for our case. We propose a novel approach to calculate a feasible scaling factor x in this section.

GEDF-VD Now, we provide an overview of our algorithm (GEDF-VD) considering an implicit-deadline sporadic MC gang task system τ to be scheduled on M identical cores. The GEDF-VD algorithm starts by checking whether GEDF can successfully schedule the *regular task system*. A regular task system denotes that, all LO-criticality tasks will execute up-to their LO-criticality WCET and all HI-criticality tasks will execute up-to their HI-criticality WCET. It returns *SUCCESS* immediately if the regular task system is schedulable. Otherwise, all HI-criticality tasks can execute up-to their LO-criticality WCETs and their deadline is shortened (i.e., virtual deadline) and set to $\hat{T}_i = xT_i$, while all LO-criticality tasks execute up-to their LO-criticality WCETs with their original deadline. If any of the currently executing job (of a HI-criticality task) executed beyond its LO-criticality WCET and did not signal its completion by \hat{T}_i , the scheduler immediately discards all currently active LO-criticality jobs. Also, the deadline for all HI-criticality jobs is changed to their release time plus their actual deadline. Section 3.2 provides a detailed description of GEDF-VD algorithm.

Algorithm 1: GEDF-VD (online part)

Input: A dual-criticality task-set $\tau = \{\tau_1, \tau_2, \dots, \tau_n\}$ and a feasible x .

```

1 /* Handling tasks at run-time */
2 Whenever a job is released by tasks  $\tau_i$  at time instant  $t$ 
3 if  $\tau_i \in \tau_{HI}$  then
4   |  $d_{i,j} = t + xT_i$ ;
5 end
6 if  $\tau_i \in \tau_{LO}$  then
7   |  $d_{i,j} = t + T_i$ ;
8 end
9 Schedule all active jobs by GEDF according to  $d_{i,j}$ 's.
10 if  $\exists \tau_{i,j} \in \tau_{HI}$  that is not finished by  $d_{i,j}$  at time  $t$  then
11   /* Mode Switch */
12   for  $\forall \tau_i \in \tau_{HI}$  do
13     |  $d_{i,j} = d_{i,j} + (1-x)T_i$ 
14   end
15   Discard all  $\tau_i \in \tau_{LO}$ 
16   Schedule  $\tau' = \{\tau_{HI}\}$  by GEDF.
17 end

```

3.2 GEDF-VD: a detailed description

In this subsection, we describe the GEDF-VD scheduling approach in a two-phase process. First, we describe what happens prior to run-time (denoted as a *pre-processing phase*). In this phase, GEDF-VD determines whether (or not) it is required to set a virtual deadline for the HI-criticality tasks. A lower and an upper bound of the virtual deadline is also calculated in this phase. Then, we discuss how the jobs are scheduled at run-time (denoted as *handling the dispatched jobs at run-time*). We present the pseudo-code for (the run-time part of) GEDF-VD in Algorithm 1.

Pre-processing phase In this phase, we perform a schedulability test for *ordinary (non-MC)* GEDF to determine whether (or not) it can successfully schedule: (i) all $\tau_i \in \tau_{lo}$ up-to their LO-criticality WCET (c_i^{lo}), and (ii) all $\tau_i \in \tau_{hi}$ up-to their HI-criticality WCET (c_i^{hi}). If the GEDF test fails, then, for each HI-criticality task $\tau_i \in \tau_{hi}$, a virtual deadline \hat{T}_i is computed (Step-2), and they execute up-to their LO-criticality WCET (c_i^{lo}).

Step 1 We start by checking whether the task-set can be successfully scheduled by GEDF. If so, then GEDF directly schedules the system. Else, we modify the task deadlines (Step 2).

Step 2 An additional *virtual deadline* parameter \hat{T}_i is calculated for each HI-criticality task τ_i , where $\hat{T}_i = xT_i$. A schedulability test for GEDF-VD is provided next. Furthermore, when the schedulability test is passed, x can be arbitrarily chosen from the range $[A, B]$ while GEDF-VD is guaranteed to generate an MC-correct schedule, where A and B are defined and can be easily calculated for any given system by the following equations:

$$A = \max\{A_1, A_2\}; \quad (1)$$

$$A_1 = \max_{i: \tau_i \in \tau_{lo}} \left\{ \frac{U_{hi}^{lo}}{M - \Delta_i^{lo} - U_{lo}^{lo}} \right\}; \tag{2}$$

$$A_2 = \max_{i: \tau_i \in \tau} \left\{ \frac{m_i^{lo} U_{hi}^{lo} + u_i^{lo} (M - \Delta_i^{lo} - m_i^{lo})}{m_i^{lo} (M - \Delta_i^{lo} - U_{lo}^{lo})} \right\}; \tag{3}$$

$$B = \min_{i: \tau_i \in \tau_{hi}} \left\{ 1 - \frac{m_i^{hi} U_{hi}^{hi} + u_i^{hi} (M - \Delta_i^{hi} - m_i^{hi})}{m_i^{hi} \times (M - \Delta_i^{hi})} \right\}. \tag{4}$$

A schedulability test for GEDF-VD The following theorem provides a sufficient schedulability test for GEDF-VD.

Theorem 1 *An MC gang task system is schedulable under GEDF-VD upon M identical unit-speed processors if both conditions hold:*

$$U_{lo}^{lo} < M - \max_i \{ \Delta_i^{lo} \}, \tag{5}$$

$$A \leq B. \tag{6}$$

We will prove this theorem later by proving Lemmas 1 and 2 in Sects. 3.3 and 3.4.

Recall that, $\Delta_i^{lo} < M$ and $\Delta_i^{hi} < M$ for all i . Therefore, $m_i U_{hi}^{lo} + u_i^{lo} (M - \Delta_i^{lo} - m_i^{lo}) = m_i^{lo} (U_{hi}^{lo} - u_i^{lo}) + u_i^{lo} (M - \Delta_i^{lo}) > 0$, which with (5) together implies $A > 0$; and also $m_i^{hi} U_{hi}^{hi} + u_i^{hi} (M - \Delta_i^{hi} - m_i^{hi}) = m_i^{hi} (U_{hi}^{hi} - u_i^{hi}) + u_i^{hi} (M - \Delta_i^{hi}) > 0$, which implies $B < 1$. Thus, both (5) and (6) being true implies that $0 < A \leq B < 1$, which guarantees that any x chosen from $[A, B]$ must be a valid scaling factor such that $0 < x < 1$.

Run-time dispatch Similar to GEDF, at any specific time instant, a task with the earliest deadline gets the highest priority. In case of ties, task with a smaller index is favored. Let a binary variable ξ indicate the system-criticality level, then consider the following two possible cases:

Case 1 System is in the LO-criticality mode ($\xi = 0$), j th job of task τ_i arrives at time t :

(i) If τ_i is a LO-criticality task, set the deadline as $d_{ij} = t + T_i$, else set $d_{ij} = t + \hat{T}_i$, where $\hat{T}_i = xT_i$.

(ii) If any of the currently executing jobs executes for more than c_i^{lo} and does not signal completion, then the system switches to the HI-criticality mode (Case 2).

Case 2 While the system is in the HI-criticality mode ($\xi = 1$):

- (i) Discard all LO-criticality tasks (or use background scheduling).
- (ii) Update the deadline for the currently active HI-criticality jobs into release time (of these jobs) plus their actual relative deadline.
- (iii) For any future HI-criticality task τ_i that releases a job at time t , the deadline is set to $t + T_i$.
- (iv) When there is an idle instant, switch to the LO-criticality mode (Case 1)¹.

3.3 Proof of correctness in the LO-criticality mode

In this subsection, we show that GEDF-VD and its schedulability test given by Theorem 1 are able to guarantee MC correctness at LO-criticality mode.

Lemma 1 *If both (5) and (6) are true, GEDF-VD guarantees that all LO-criticality tasks meet their deadlines and all HI-criticality tasks meet their virtual deadlines during LO-criticality mode.*

Proof Dong and Liu (2017) have proved that, given any real-time implicit deadline sporadic gang task system τ , GEDF can schedule it successfully if

$$\begin{aligned} U_{sum} &\leq (M - \Delta_i) \times \left(1 - \frac{u_i}{m_i}\right) + u_i \\ \Leftrightarrow U_{sum} &\leq M - \Delta_i + u_i \left(1 - \frac{M - \Delta_i}{m_i}\right) \end{aligned} \quad (7)$$

holds for all $\tau_i \in \tau$ (refer to Theorem 2). The virtual deadline increases the utilization of these HI-criticality tasks (and hence the whole system). Note that, in the LO-criticality mode, each HI-criticality task is scheduled by its virtual relative deadline xT_i while each LO-criticality task is scheduled by its actual deadline T_i . Therefore, it is sufficient to view each LO-criticality task as a sporadic task with utilization u_i^{lo} and view each HI-criticality task as a sporadic task with utilization u_i^{lo}/x , in order to meet every LO-criticality deadline and every HI-criticality virtual deadline in LO-criticality mode. Then, for every i such that $\tau_i \in \tau$, we discuss the two cases for $M - \Delta_i^{lo} - m_i^{lo}$. Therefore, it suffice to evaluate (7) under such utilizations for every task τ_i . We show this by two cases: (1) $\tau_i \in \tau_{hi}$, and (2) $\tau_i \in \tau_{lo}$.

Case 1 $\tau_i \in \tau_{hi}$. In this case, using (7) as a result from Dong and Liu (2017), we just need the following inequality to hold for any $\tau_i \in \tau_{hi}$.

¹ Note that HI-criticality mode exists for certification purposes. Such both directions of mode switch should be unlikely events during run time. Please also refer to the discussions about apriori verification and run-time robustness in Sect. 1.

$$\begin{aligned}
 U_{lo}^{lo} + \frac{U_{hi}^{lo}}{x} &\leq M - \Delta_i^{lo} + \frac{u_i^{lo}}{x} \left(1 - \frac{M - \Delta_i^{lo}}{m_i^{lo}} \right) \\
 \Leftrightarrow \frac{U_{hi}^{lo}}{x} + \frac{u_i^{lo}}{x} \left(\frac{M - \Delta_i^{lo}}{m_i^{lo}} - 1 \right) &\leq M - \Delta_i^{lo} - U_{lo}^{lo} \\
 \Leftrightarrow \frac{m_i^{lo} U_{hi}^{lo} + u_i^{lo} (M - \Delta_i^{lo} - m_i^{lo})}{m_i^{lo} \cdot x} &\leq M - \Delta_i^{lo} - U_{lo}^{lo}
 \end{aligned}
 \tag{8}$$

Notice that (5) implies

$$M - \Delta_i^{lo} - U_{lo}^{lo} > 0 \text{ for all } i \text{ such that } \tau_i \in \tau, \tag{9}$$

and (6) allows $x \in [A, B]$ can be chosen so that $x \geq A$, which, by (1) and (3), implies

$$x \geq \frac{m_i^{lo} U_{hi}^{lo} + u_i^{lo} (M - \Delta_i^{lo} - m_i^{lo})}{m_i^{lo} (M - \Delta_i^{lo} - U_{lo}^{lo})} \text{ for all } i \text{ such that } \tau_i \in \tau. \tag{10}$$

It is clear that (9) and (10) imply (8).

Case 2 $\tau_i \in \tau_{lo}$. In this case, using (7) as a result from Dong and Liu (2017), we just need the following condition to hold for any $\tau_i \in \tau_{lo}$.

$$U_{lo}^{lo} + \frac{U_{hi}^{lo}}{x} \leq M - \Delta_i^{lo} + u_i^{lo} \left(1 - \frac{M - \Delta_i^{lo}}{m_i^{lo}} \right) \tag{11}$$

Subcase 2.1 $M - \Delta_i^{lo} - m_i^{lo} \leq 0$. In this case, $M - \Delta_i^{lo} \leq m_i^{lo} \implies 1 - \frac{M - \Delta_i^{lo}}{m_i^{lo}} \geq 0$. Therefore, the following inequality implies (11):

$$U_{lo}^{lo} + \frac{U_{hi}^{lo}}{x} \leq M - \Delta_i^{lo}. \tag{12}$$

Notice that (5) implies

$$M - \Delta_i^{lo} - U_{lo}^{lo} > 0 \text{ for all } i \text{ such that } \tau_i \in \tau, \tag{13}$$

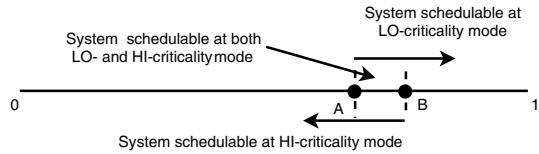
and (6) allows $x \in [A, B]$ can be chosen so that $x \geq A$, which, by (1) and (2), implies

$$x \geq \frac{U_{hi}^{lo}}{M - \Delta_i^{lo} - U_{lo}^{lo}} \text{ for all } i \text{ such that } \tau_i \in \tau_{lo}. \tag{14}$$

It is clear that (13) and (14) imply (12)

Subcase 2.2 $M - \Delta_i^{lo} - m_i^{lo} > 0$. In this case, $M - \Delta_i^{lo} > m_i^{lo} \implies 1 - \frac{M - \Delta_i^{lo}}{m_i^{lo}} < 0$. So, $\frac{u_i^{lo}}{x} \left(1 - \frac{M - \Delta_i^{lo}}{m_i^{lo}} \right) < u_i^{lo} \left(1 - \frac{M - \Delta_i^{lo}}{m_i^{lo}} \right)$, as $0 < x < 1$. Therefore, the following inequality implies (11).

Fig. 2 Any value of the scaling factor x , where $A \leq x \leq B$, guarantees an MC-correct schedule



$$U_{lo}^{lo} + \frac{U_{hi}^{lo}}{x} \leq M - \Delta_i^{lo} + \frac{u_i^{lo}}{x} \left(1 - \frac{M - \Delta_i^{lo}}{m_i^{lo}}\right) \tag{15}$$

By the same reasoning as that for Case 1, (15) always holds because (9) and (10) are “for any $\tau_i \in \tau$ ” and both HI- and LO-criticality tasks are included in the set τ . That is, (11) is also true in Case 2.2 here.

Combining Cases 1 and 2 (the latter includes Sub-cases 2.1 and 2.2), the lemma follows. □

3.4 Proof of correctness in the HI-criticality mode

In this subsection, we show that GEDF-VD and its schedulability test given by Theorem 1 are able to guarantee MC correctness at HI-criticality mode.

Lemma 2 *If both (5) and (6) are true, GEDF-VD guarantees that all HI-criticality tasks meet their deadlines during HI-criticality mode.*

Proof At the mode switch point from the lo- to HI-criticality mode, a job from any task $\tau_i \in \tau_{hi}$ must be either completed or has a deadline at least $(1 - x)T_i$ after this mode-switch point; otherwise, an earlier time instant would have been the mode switch point.

Afterwards, any job from any task $\tau_i \in \tau_{hi}$ has at least T_i time units (which is more than $(1 - x)T_i$ as $0 < x < 1$) from their releases in the HI-criticality mode to their corresponding deadlines.

Therefore, viewing each task $\tau_i \in \tau_{hi}$ in the HI-criticality mode as a sporadic task with utilization $\frac{u_i^{hi}}{(1-x)}$ and using (7) as a result from Dong and Liu (2017), the following inequality is sufficient to ensure that all HI-criticality tasks meet their actual deadlines during HI-criticality mode. For all i such that $\tau_i \in \tau_{hi}$,

$$m_i^{hi} \times \frac{U_{hi}^{hi}}{(1-x)} \leq m_i^{hi} \times (M - \Delta_i^{hi}) - \frac{u_i^{hi}}{1-x} \times (M - \Delta_i^{hi} - m_i^{hi}) \tag{16}$$

Notice that (6) allows $x \in [A, B]$ can be chosen so that $x \leq B$, which, by (4), implies the following equation holds for all i such that $\tau_i \in \tau_{hi}$:

$$x \leq 1 - \frac{m_i^{\text{hi}} U_{\text{hi}}^{\text{hi}} + u_i^{\text{hi}} (M - \Delta_i^{\text{hi}} - m_i^{\text{hi}})}{m_i^{\text{hi}} \times (M - \Delta_i^{\text{hi}})} \quad (17)$$

Furthermore, Equation (17) is equivalent to Equation (16), as $0 < x < 1$ and $\Delta_i^{\text{hi}} < M$. Thus, the lemma follows. \square

Finishing up We establish Theorem 1 by combining Lemma 1 and 2, and it serves as a sufficient schedulability test for GEDF-VD to schedule MC gang task sets on M identical processors. In addition, Fig. 2 gives a high-level intuition for validating Theorem 1, given that Lemma 1 and 2 have been proven. Note that we did leverage some insights (in our analysis) from prior works on MC scheduling and gang scheduling. However, considering both of these directions brought increased complexity in our system model. The existing analysis of MC scheduling or gang scheduling is not directly applicable to our work. For example, in the speedup bound analysis for MC scheduling of ordinary sporadic tasks, *an individual task's utilization is at most the speed of a processor* is a straightforward and necessary feasibility condition, while it no longer holds for the gang tasks.

Remark 2 If we consider that the degree of parallelism of a task τ_i remains unchanged even after a mode-switch, i.e., $m_i^{\text{lo}} = m_i^{\text{hi}} = m_i$, then the correctness proofs become identical to the correctness proofs established in Bhuiyan et al. (2019b).

4 Speedup bound analysis

In this section, we evaluate the effectiveness of our algorithm GEDF-VD based on *speedup bound* metric, which is a widely accepted tool for evaluating the effectiveness of multiprocessor scheduling algorithms (Baruah et al. 2015a). We will first provide the related definition and some existing results, and then (in Sect. 4.1) will derive the speedup bound for gang tasks under GEDF algorithm considering the *non-MC* systems. This is the first speedup bound for (non-MC) gang tasks under GEDF scheduling policy, which lays the foundation for deriving a speedup bound for MC gang tasks. Finally, in Sect. 4.2, considering the *MC* sporadic gang tasks, we prove a speedup bound for our proposed algorithm GEDF-VD. We derive the speedup bound for non-MC gang tasks (Sect. 4.1) and the MC gang tasks (Sect. 4.2) assuming that the degree of parallelism of a task τ_i does not change after a mode-switch, i.e., $m_i^{\text{lo}} = m_i^{\text{hi}} = m_i$.

Definition 5 (*Speedup factor and speedup bound*) For a scheduler \mathcal{S} , a speedup factor \mathcal{V} ($\mathcal{V} \geq 1$) (also known as resource augmentation factor) means that any task set that is schedulable by an optimal scheduler on a platform of speed-1 cores will be schedulable by \mathcal{S} on a platform of speed- \mathcal{V} cores.

For a scheduler \mathcal{S} , a speedup bound refers to the lower bound of the speedup factor \mathcal{V} achievable by it. A speedup bound for a scheduler \mathcal{S} provides an estimation of how far the performance of \mathcal{S} is from an optimal scheduler, and the lower the better.

Limitations Our speedup factors result in this section rely on the following assumption that

$$m_i \leq \frac{M+1}{2} \text{ for all } \tau_i \in \tau, \quad (18)$$

The speedup factors results in this section apply only to systems that satisfy the condition (18). Nonetheless, condition (18) was not required for the schedulability test and analysis in the last section. Therefore those schedulability results apply to a broader range of MC gang task systems. In practice, condition (18) is often satisfied because of the number of cores in modern platforms increases.

Note that gang tasks cannot be scheduled on uniprocessor platforms due to their natures of the *mandatory* parallel processor access request. Therefore, in order to compare with a potential optimal scheduler on a uniprocessor, we propose a *De-gang* transformation between a multiprocessor gang task set and a corresponding Liu-and-Layland (LL) task set:

- **De-gang** Given a gang task set $\tau = \{\tau_1, \dots, \tau_n\}$, for each task $\tau_i = \{m_i, c_i, T_i\}$, construct m_i LL tasks $\{\tau_i^{(1)}, \dots, \tau_i^{(m_i)}\}$, each with the same execution length and period, i.e., $\tau_i^{(j)} = \{c_i, T_i\}$ for any $j = 1, \dots, m_i$. For mapping of the other way around, any deganged LL task set can be clustered into n groups, where there are m_i tasks from the i th group sharing the same execution time c_i and the same period T_i , resulting in a gang task $\tau_i = \{m_i, c_i, T_i\}$ of the same “total” utilization. The extension to MC task set is trivial—treat c_i as a vector and maintain the values during the transfer.

A moment thought should convince the reader that it suffices to restrict our attention to the de-ganged LL task set when deriving the speedup bound, as the de-ganged LL task set being schedulable is *necessary* for the corresponding gang task set to be schedulable. This transformation does not change the overall set utilization and thus does not change the utilization-based necessary schedulability conditions (i.e., basis of the speedup proofs). Throughout the proofs in this section, the following Greek letters will be used frequently:

$$\begin{aligned} \psi &= M / \left(2 - \frac{1}{M}\right); \\ \phi &= \frac{\sqrt{5} + 1}{2} \text{ (i.e., golden ratio);} \\ \Phi &= \frac{\sqrt{5} - 1}{2} \end{aligned} \quad (19)$$

4.1 Speedup bound for gang tasks under GEDF

In this subsection, we derive the speedup bound (shown in Theorem 2) for the algorithm GEDF, considering the *non-MC gang task set* τ , executing on \mathcal{V} -speed cores. This is the first speedup bound result for gang task under GEDF scheduling. This analysis lays the basis for deriving the speedup bound for the proposed MC gang task scheduler.

Theorem 2 *Given any de-ganged task set that is schedulable on a speed-M uni-processor, the corresponding gang task set will pass the schedulability test of GEDF upon a M-core system, each of speed $\mathcal{V} = 2 - 1/(M + 1 - \min_i\{m_i\})$.*

Proof Because $1 \leq m_i \leq M$ for any i , we know that for all $\tau_i \in \tau$,

$$\begin{aligned} \mathcal{V} &= 2 - \frac{1}{M + 1 - \min_i\{m_i\}} \\ &\geq 2 - \frac{1}{M + 1 - m_i} \\ &= \frac{2M + 1 - 2m_i}{M + 1 - m_i}; \end{aligned} \tag{20}$$

From feasibility of the LL task set on a speed-M uniprocessor, we have $U_{sum} \leq M$. So,

$$\begin{aligned} (20) \implies \mathcal{V} &\geq \frac{U_{sum} + M + 1 - 2m_i}{M + 1 - m_i} \\ &\iff \frac{U_{sum}}{\mathcal{V}} \leq M - (m_i - 1) + \frac{(2m_i - M - 1)}{\mathcal{V}} \\ &\iff m_i \frac{U_{sum}}{\mathcal{V}} \leq m_i M - m_i(m_i - 1) + \frac{m_i}{\mathcal{V}}(2m_i - M - 1) \\ &\implies m_i \frac{U_{sum}}{\mathcal{V}} \leq m_i M - m_i(m_i - 1) + \frac{u_i}{\mathcal{V}}(2m_i - M - 1) \\ &\qquad [u_i \leq m_i, 2m_i - M - 1 \leq 0] \end{aligned} \tag{21}$$

The condition $2m_i - M - 1 \leq 0$ is equivalent to (18); while $u_i \leq m_i$ is true for any gang task because the utilization of each gang task τ_i is $u_i = m_i(c_i/T_i)$, where $c_i \leq T_i$. Note that u_i can also be viewed as the total utilization of the m_i de-ganged LL tasks that correspond to the gang task τ_i . Again, de-ganging preserves the utilization of the set. From Eq. (21):

$$\begin{aligned}
m_i \frac{U_{sum}}{\mathcal{V}} &\leq m_i M - (m_i - 1)m_i + (m_i - 1) \frac{u_i}{\mathcal{V}} + \frac{u_i}{\mathcal{V}}(m_i - M) \\
&\implies m_i \frac{U_{sum}}{\mathcal{V}} \leq m_i M - (m_i - 1) \left(m_i - \frac{u_i}{\mathcal{V}} \right) - \frac{u_i}{\mathcal{V}}(M - m_i) \\
&\implies m_i \frac{U_{sum}}{\mathcal{V}} \leq m_i M - \Delta_i \left(m_i - \frac{u_i}{\mathcal{V}} \right) - \frac{u_i}{\mathcal{V}}(M - m_i)
\end{aligned}$$

[From Definition 4: $0 \leq \Delta_i \leq m_i - 1$]

$$\begin{aligned}
&= m_i(M - \Delta_i) - (M - \Delta_i - m_i) \frac{u_i}{\mathcal{V}} \quad [\text{re-arrange}] \\
&= (M - \Delta_i) \left(m_i - \frac{u_i}{\mathcal{V}} \right) + m_i \frac{u_i}{\mathcal{V}} \quad [\text{re-arrange}] \\
\implies \frac{U_{sum}}{\mathcal{V}} &\leq (M - \Delta_i) \left(1 - \frac{u_i}{m_i \mathcal{V}} \right) + \frac{u_i}{\mathcal{V}} \\
&\quad [\text{dividem}_i \text{ on both sides, re-arrange}], \text{ for all } i
\end{aligned} \tag{22}$$

The equation above implies that the corresponding gang set is GEDF schedulable on M speed- \mathcal{V} processors (Theorem 2 in Dong and Liu (2017)). Note that the last step is true because under speed of \mathcal{V} , all utilizations in the test should be treated as the ones under speed 1 divided by \mathcal{V} , in order to apply the original schedulability test under a speed-1 platform. \square

Theorem 2 indicates that the speedup factor of the GEDF schedulability test in Theorem 2 of Dong and Liu (2017) (for gang task set) is no greater than $\mathcal{V} = 2 - 1/(M + 1 - \min_i \{m_i\})$. Because $1 \leq m_i \leq M$ for any i , $\mathcal{V} \leq 2 - \frac{1}{M}$. Therefore, the following corollary follows directly from Theorem 2.

Corollary 1 *Given any de-ganged task set that is schedulable on a speed- M uni-processor, the corresponding gang task set will pass the schedulability test of GEDF upon a M -core system, each of speed $(2 - \frac{1}{M})$.*

Furthermore, scaling all speeds by a factor of $1/(2 - \frac{1}{M})$ lead to the following corollary.

Corollary 2 *Given any de-ganged task set that is schedulable on a speed- ψ uni-processor, the corresponding gang task set will pass the schedulability test of GEDF upon M unit-speed processors, where $\psi = M/(2 - \frac{1}{M})$.*

4.2 Speedup bound for gang tasks under GEDF-VD

The previous subsection proved the speedup bound for non-MC task under GEDF. We now brings MC and virtual deadlines into the picture, and derive the speedup

bound for *MC gang task set* τ under GEDF-VD. From the definitions of ϕ and Φ in Equation (19), the following properties hold:

$$1 + \Phi = \phi = \frac{1}{\Phi} \tag{23}$$

$$\Phi + \Phi^2 = 1 \tag{24}$$

Theorem 3 *Given any de-ganged MC task set that is schedulable on a speed- $(\psi \cdot \Phi)$ uniprocessor, the corresponding MC gang task set will be schedulable under GEDF-VD upon M unit-speed processors, where $\psi = M/(2 - \frac{1}{M})$ and $\Phi = \frac{\sqrt{5}-1}{2}$.*

Proof The de-ganged MC task set being schedulable on a speed- $(\psi \cdot \Phi)$ uni-processor implies

$$\max\{U_{lo}^{lo} + U_{hi}^{lo}, U_{hi}^{hi}\} \leq \psi \cdot \Phi. \tag{25}$$

We proceed the rest of this proof in two cases.

Case 1: $U_{hi}^{lo} \geq \Phi \cdot U_{lo}^{lo}$. By (25) and the condition of Case 1,

$$\begin{aligned} \psi \cdot \Phi &\geq U_{lo}^{lo} + U_{hi}^{lo} \geq (1 + \Phi)U_{lo}^{lo} \\ \implies U_{lo}^{lo} &\leq \frac{\Phi}{1 + \Phi} \cdot \psi = \Phi^2 \cdot \psi. \text{ [by (23)]} \end{aligned}$$

Then, by the above and (25),

$$U_{lo}^{lo} + U_{hi}^{hi} \leq \Phi^2 \cdot \psi + \Phi \cdot \psi = \psi. \text{ [by (24)]}$$

Thus, no virtual deadline needs to be set at all. Both HI- and LO-criticality tasks are scheduled by GEDF according to their actual deadlines on M unit-speed processors. By Corollary 2, no deadline will be missed.

Case 2: $U_{hi}^{lo} < \Phi \cdot U_{lo}^{lo}$. By (25) and the condition of Case 2,

$$\begin{aligned} \psi \cdot \Phi &\geq U_{lo}^{lo} + U_{hi}^{lo} > (\frac{1}{\Phi} + 1)U_{hi}^{lo} \\ &= \frac{1 + \Phi}{\Phi} \cdot U_{hi}^{lo} = \frac{1}{\Phi^2} \cdot U_{hi}^{lo}. \text{ [by (23)]} \end{aligned}$$

That is,

$$U_{hi}^{lo} < \Phi^3 \cdot \psi. \tag{26}$$

Then, we have

$$\begin{aligned}
 U_{lo}^{lo} + \frac{U_{hi}^{lo}}{1 - U_{hi}^{hi}/\psi} &= U_{lo}^{lo} + U_{hi}^{lo} + U_{hi}^{lo} \cdot \frac{U_{hi}^{hi}/\psi}{1 - U_{hi}^{hi}/\psi} \\
 &\leq U_{lo}^{lo} + U_{hi}^{lo} + U_{hi}^{lo} \cdot \frac{\Phi}{1 - \Phi} \\
 &\quad [U_{hi}^{hi} \leq \psi \cdot \Phi \text{ by (25)}] \\
 &= U_{lo}^{lo} + U_{hi}^{lo} + U_{hi}^{lo} \cdot \frac{\Phi}{\Phi^2} \text{ [by (24)]} \\
 &< \psi \cdot \Phi + \Phi^3 \cdot \psi \cdot \frac{\Phi}{\Phi^2} \text{ [by (25) and (26)]} \\
 &= (\Phi + \Phi^2)\psi \text{ [rearrange]} = \psi, \text{ [by (24)]}
 \end{aligned}$$

which is concluded as

$$U_{lo}^{lo} + \frac{U_{hi}^{lo}}{1 - U_{hi}^{hi}/\psi} < \psi. \tag{27}$$

In this case, one could take $x = \frac{U_{hi}^{lo}}{\psi - U_{lo}^{lo}}$ as the scaling factor to set the virtual deadlines for HI-criticality tasks. Because the de-ganged task set is schedulable on a speed- $(\psi \cdot \Phi)$ uniprocessor, $U_{lo}^{lo} \leq \psi \cdot \Phi$, which implies $x > 0$, as $\Phi < 1$, $\psi > 0$, and $U_{hi}^{lo} > 0$. On the other hand, $U_{hi}^{lo} < \Phi \cdot U_{lo}^{lo}$ in Case 2, so

$$\begin{aligned}
 x &= \frac{U_{hi}^{lo}}{\psi - U_{lo}^{lo}} < \frac{\Phi \cdot U_{lo}^{lo}}{\psi - U_{lo}^{lo}} \\
 &\leq \frac{\Phi \cdot \psi \cdot \Phi}{\psi - \psi \cdot \Phi} = \frac{\Phi^2}{1 - \Phi} = 1. \text{ [by (24)]}
 \end{aligned}$$

Thus, in this case, $0 < x < 1$ is guaranteed under this particular setting and therefore this x can always be used as the scaling factor to set the virtual deadlines for GEDF-VD.

Then, we first show that all LO-criticality tasks meet their actual deadlines and all HI-criticality tasks meet their *virtual* deadlines during the LO-criticality mode.

$$U_{lo}^{lo} + \frac{U_{hi}^{lo}}{x} = U_{lo}^{lo} + \psi - U_{lo}^{lo} = \psi.$$

By Corollary 2, the above equation implies that using GEDF-VD to schedule the gang task set on M unit-speed processors, all LO-tasks meet their actual deadlines and all HI-tasks meet their *virtual* deadlines during the LO-mode. Next, we show that all HI-criticality tasks, including any carryover (HI-) jobs across the mode-switch point, meet their *actual* deadlines during the HI-criticality mode.

Because the virtual deadlines are set as $x \cdot T_i$ for each HI-criticality task τ_i , every HI-criticality job including the one triggering the mode switch will have at least $(1 - x)T_i$ time units to finish its at most C_i^{hi} execution and to release its next job. It suffices to consider the schedulability when replacing each HI-criticality task in

the HI-criticality mode by a implicit-deadline sporadic task with period $(1 - x)T_i$ and execution C_i^{hi} . It can be done by checking their total utilization

$$\sum_{\tau_i \in \tau_{hi}} \frac{C_i^{hi}}{(1 - x)T_i} = \frac{U_{hi}^{hi}}{1 - x}.$$

On the other hand, by (27), we have

$$\frac{U_{hi}^{lo}}{\psi - U_{lo}^{lo}} < 1 - U_{hi}^{hi} / \psi,$$

and $1 - x > U_{hi}^{hi} / \psi$ holds since we set $x = \frac{U_{hi}^{lo}}{\psi - U_{lo}^{lo}}$. Thus,

$$\frac{U_{hi}^{hi}}{1 - x} < \frac{U_{hi}^{hi}}{U_{hi}^{hi} / \psi} = \psi.$$

By Corollary 2, the above equation implies that using GEDF-VD to schedule the gang task set on M unit-speed processors, all HI-criticality tasks, including any carryover (HI-) jobs across the mode-switch point, meet their *actual* deadlines during the HI-criticality mode. □

Finally, we can easily use Theorem 3 to derive a speedup bound for GEDF-VD to schedule MC gang task sets on identical processors, as stated in the following theorem.

Theorem 4 *If any potentially optimal algorithm can schedule an MC gang task set on M unit-speed processors, GEDF-VD is able to schedule the same MC gang task set on M speed- $(\sqrt{5} + 1)$ processors.*

Proof Theorem 3 directly implies that:

If any potentially optimal algorithm can schedule an MC gang task set on M speed- $(\psi \cdot \Phi / M)$ processors, GEDF-VD is able to schedule the same MC gang task set on M unit-speed processors.

This is because for a MC gang task set to be schedulable on M speed- $(\psi \cdot \Phi / M)$ processors, it is necessary for its corresponding de-ganged MC task set to schedulable on a speed- $(\psi \cdot \Phi)$ uniprocessor. Note that, by definitions: $\psi = \frac{M}{2 - \frac{1}{M}}$ and $\Phi = \frac{\sqrt{5}-1}{2}$, the following statement is true:

If any potentially optimal algorithm can schedule a MC gang task set on M speed- $(\frac{1}{2 - \frac{1}{M}} \cdot \frac{\sqrt{5}-1}{2})$ processors, GEDF-VD is able to schedule the same MC gang task set on M unit-speed processors.

Scaling the speed unit up by $(2 - \frac{1}{M}) \frac{\sqrt{5}+1}{2}$ (please note that $\frac{\sqrt{5}-1}{2} \cdot \frac{\sqrt{5}+1}{2} = 1$), the above statement can be re-written as:

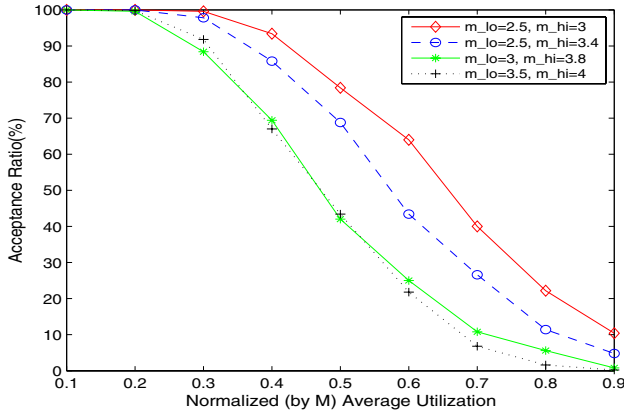


Fig. 3 Acceptance ratio for GEDF-VD with a different (after a mode-switch) average degrees of parallelism

Table 2 Acceptance ratio for different amount of tasks generated under various average utilization and R value

$U_{avg} \rightarrow$ # of tasks↓	R	2	2.5	3	3.5	4	4.5	5
8	4	100	100	97	59	5	5	4
12	4	100	99	94	65	2	2	2
16	4	100	100	98	50	0	0	0
8	8	97	75	67	31	23	19	11
12	8	100	100	74	58	40	17	6
16	8	100	98	91	77	48	34	11

If any potentially optimal algorithm can schedule a MC gang task set on M unit-speed processors, GEDF-VD is able to schedule the same MC gang task set on M speed- $\left(2 - \frac{1}{M}\right) \frac{\sqrt{5}+1}{2}$ processors.

Since $\left(2 - \frac{1}{M}\right) \frac{\sqrt{5}+1}{2} < \sqrt{5} + 1$, the theorem follows. □

5 Evaluation

In this section, we evaluate the performance of GEDF-VD through simulation results. While the simulation results provide some representation of the proposed scheduling’s performance, they may not represent the exact behavior of our proposed approach in real systems for several reasons. For example, memory plays a vital role from an implementation point of view and needs to be available and allocated to parallel threads. Although a recent work (Capodieci et al. 2018) has implemented a preemptive EDF scheduler for GPU tasks providing bandwidth isolation, MC scheduling on the GPU platform (with preemptive EDF) and the memory

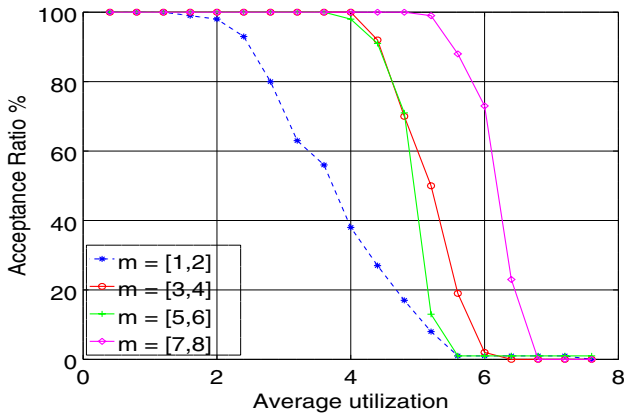


Fig. 4 Acceptance ratio for GEDF-VD in an 8-core platform with $R = 4$, and under same ranges of degrees of parallelism

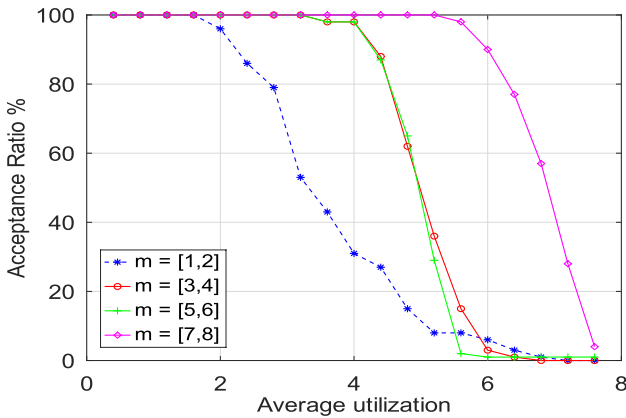


Fig. 5 Acceptance ratio for GEDF-VD in an 8-core platform with $R = 8$, and under same ranges of degrees of parallelism

partitioning technique to the gang tasks are yet to be explored. In the future, we plant to explore implementation and experimentation on a real hardware platform. As our work is the first to propose MC gang task scheduling, there is no perfect baseline for comparison. We have performed many experiments by varying different factors to observe the efficiency of our algorithm.

5.1 Experimental setup

Workload generation We generate MC gang tasks based on the following parameters.

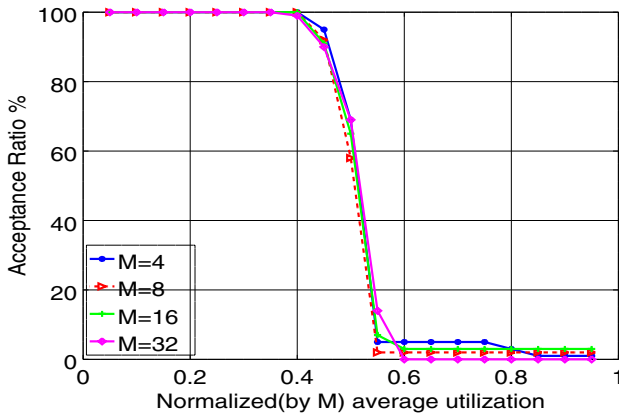


Fig. 6 Acceptance ratio for GEDF-VD in an M -core platform (with $R = 4$). In this experiment $M \in \{4, 8, 16, 32\}$

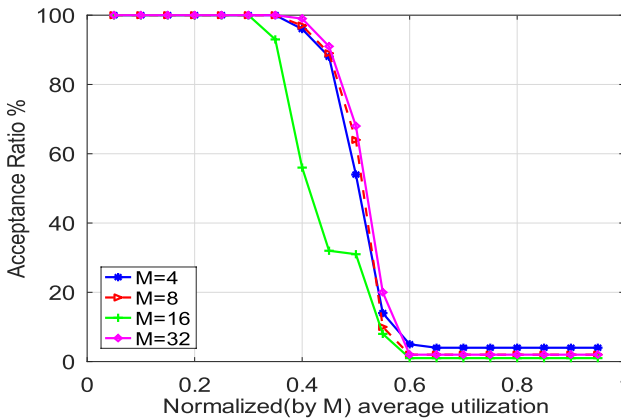


Fig. 7 Acceptance ratio for GEDF-VD in an M -core platform (with $R = 8$). In this experiment $M \in \{4, 8, 16, 32\}$

- M : The number of processor cores.
- $m_{min}, m_{max}, m_{avg}$: The minimum, maximum, and average value for m (i.e., degree of parallelism), respectively. We generate the task set by varying these three parameters, where $m_{min}, m_{max} \in [1, M]$ and $m_{min} \leq m_{avg} \leq m_{max}$.
- U_{avg} : The average utilization for the task set. We have varied U_{avg} value from $0.05 \times M$ to $0.95 \times M$ with $0.05 \times M$ difference at each step.
- $P_{hi} = 0.5$: The probability of a task $\tau_i \in \tau_{hi}$.
- R : Denotes the maximum ratio of u_i^{hi} to u_i^{lo} , where $R \in [4, 8]$. We generate u_i^{hi} uniformly from $[u_i^{lo}, R \times u_i^{lo}]$.

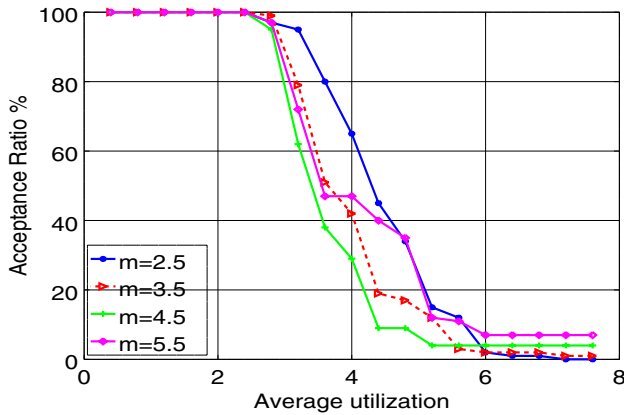


Fig. 8 Acceptance ratio for GEDF-VD in an 8-core platform with $R = 4$ and a varying range of m_{avg}

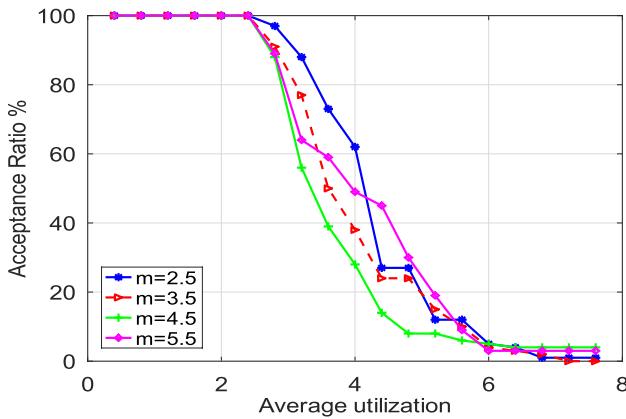


Fig. 9 Acceptance ratio for GEDF-VD in an 8-core platform with $R = 8$ and a varying range of m_{avg}

At first, for a specific value of n (number of tasks per task set), we generate the m values for each task. m is uniformly generated from $[m_{min}, m_{max}]$ range in a way so that the average m for all tasks remains equal to m_{avg} . Next, for a specific value of average utilization U_{avg} , we calculate the average utilization u_i^a for each task by following the log-normal distribution. Note that, for n number of gang tasks, there are total $\sum_{i=1}^n m_i = m_{avg} \times n$ amount of single task instances in each task set. For the sake of a proper distribution, we extend the *UUniFast algorithm* (Bolado et al. 2004) for Gang task. We use log-normal distribution over $\sum_{i=1}^n m_i$ task instances similarly as UUniFast, but for a single task, we take the average of all of its instances as the task’s average utilization. The values of u_i^{lo} is uniformly generated from $[\frac{2 \times u_i^a}{R+1}, u_i^a]$ so that the value of u_i^{hi} is always in the range $[u_i^{lo}, R \times u_i^{lo}]$.

Simulation setup We performed the simulation for average utilization ranging from $0.05M$ to $0.95M$ with a step size of $0.05M$. For each average, 100 task sets (each with 10 tasks) are generated.

5.2 Evaluation results

We execute a set of gang tasks under our proposed algorithm by varying different parameters, and present the simulation results in Figs. 3, 4, 5, 6, 7, 8 and 9, and in Table 2. Note that, in Table 2, and in Figs. 4, 5, 6, 7, 8 and 9, we consider that the degree of parallelism (of the HI-criticality tasks) does not change after a mode-switch.

Effect of changing the degree of parallelism after a mode switch In this experiment, we set the value for M to 8, and vary (i.e., increase or unchanged) the degree of parallelism of a HI-criticality task after a mode-switch. To incorporate the change in the task model, i.e., the degree of parallelism of a HI-criticality task can change after a mode-switch, we slightly modify the simulation setup described in Sect. 5.1. We report the acceptance ratio (in Fig. 3) with a different average of m_i^{lo} (i.e., from 2.5 to 3.5) and m_i^{hi} (i.e., from 3 to 4) values. This figure reports that the *acceptance ratio* (i.e., the ratio of the number of schedulable task sets over the total number of task sets) decreases when the degree of parallelism increases, which can be explained by Eqs. (8) and (16). That is, a higher value of m_i^{lo} or m_i^{hi} inversely affects the acceptance ratio.

Effect of changing the degree of parallelism in a range with lower difference In this set of experiments, for $M = 8$, we vary a task's degree of parallelism (m) in a different range, while the difference between the upper and lower bound in each range is fixed. The acceptance ratio under varying degree of parallelism (and different R values) is reported in Figs. 4 and 5. These figures indicate that in boundary cases (where the degree of parallelism is very low or very high) acceptance ratio changes proportionally with respect to the degree of parallelism. This behavior can be explained with the help of Eqs. (8) and (16). When m increases or decreases by a large amount, acceptance ratio will increase or decrease respectively. However, for a small change of m , acceptance ratio may not change proportionally. This is because the schedulability conditions provided by Eqs. (8) and (16) are also effected by the maximum number of idle cores (Δ_i), which is dependent on m .

Effect of changing the total number of cores In Figs. 6 and 7, we report the acceptance ratio of the task set by varying the number of cores in the system, M . In this set of experiments, we set a value for m_{avg} , which is uniformly generated from a range of $[\frac{M}{2}, \frac{3M}{4}]$. Simulations are conducted for $M = 4, 8, 16$, and 32 and the average utilization is weighted with respect to the value of M . Figures 6 and 7 shows that the acceptance ratio is not affected by different values of M and remains almost unchanged.

Effect of changing number of tasks per task set In this set of experiments, we have randomly generated 100 task sets with different R values and 8, 12, and 16 tasks per task set (with U_{avg} changing from 2 to 5 with a step size of 0.5) and report the acceptance ratio in Table 2. From the reported data, it is clear that the acceptance ratio of the task set is not affected by the number of tasks per set. This result indicates the effectiveness of our proposed algorithm under a varying number of tasks in a task set.

Effect of changing m_{avg} value In Figures 8 and 9, we show the acceptance ratio by varying m_{avg} in an 8-core platform. The result does not demonstrate a direct relationship between m_{avg} and the acceptance ratio.

6 Related work

Since Vestal's proposal (Vestal 2007) of MC workload model, much work has focused on scheduling MC tasks (refer to Burns and Davis (2013) for a survey). For uniprocessor platforms, many algorithms were proposed based on both fixed priority (e.g., Li and Baruah 2010; Baruah et al. 2011) and dynamic priority scheduling (e.g., Easwaran 2013). The work in Bhuiyan et al. (2019a, 2020b) proposed the precise scheduling policy, where all LO-criticality tasks receive a full-service guarantee even after a mode switch. On the other hand, numerous MC scheduling algorithms were proposed for multiprocessor platforms (Lee et al. 2014; Baruah et al. 2015b; Tobuschat and Ernst 2017; Awan et al. 2017; Trüb et al. 2017). Considering the multiprocessor platforms, Lee et al. (2014) and Baruah et al. (2015b) proposed fluid-based MC models, and a semi-partitioned based scheme is proposed by Awan et al. (2017).

Considering different parallel tasks models [e.g., synchronous task model (Andersson and de Niz 2012), DAG model (Bonifaci et al. 2013; Li et al. 2014; Guo et al. 2017; Bhuiyan et al. 2018, 2020a; Guo et al. 2019) and gang models (Kato and Ishikawa 2009; Goossens and Richard 2016; Dong and Liu 2017)] there have been a number of works that have provided the energy efficiency technique, schedulability analysis, and the speedup bound (i.e., resource augmentation bound) for various scheduling strategies. For synchronous tasks under GEDF scheduling, Andersson and de Niz (2012) proved a resource augmentation bound of 2 with constrained deadlines tasks. Considering DAG tasks (with arbitrary deadlines) under GEDF, Li et al. (2013) and Bonifaci et al. (2013) simultaneously proved a resource augmentation bound of 2. Bonifaci et al. (2013) also showed the bound to be 3 under global rate-monotonic scheduling. For implicit deadline DAG tasks under federated scheduling, a resource augmentation bound of 2 is showed by Li et al. (2014). Gang scheduling and Coscheduling was initially introduced to perform parallel processing with fine-grained interactions efficiently (Feitelson and Rudolph 1992; Ousterhout 1982; Gehringer et al. 1987). Both of these approaches allocate resources to the threads of the same task concurrently. However, gang scheduling imposes a strict requirement of executing all threads of the same task simultaneously. In contrast, in coscheduling, some threads may not execute concurrently with the rest of the threads in the same task. Some recent work used this concept to execute the parallel workload in cloud computing (Stavrinides and Karatza 2016) and extended to incorporate hard real-time tasks (Goossens and Richard 2016). The work in Goossens and Richard (2016) also has proposed a DP-Fair based scheduling of periodic gang tasks and proved a speedup bound which is no larger than $(2 - 1/m)$. A recent work by Alahmad and Gopalakrishnan (2019) proposed the isochronous scheduling, which has some similarity to the traditional gang scheduling. Unlike the gang scheduling, the isochronous model assumed that the job versions might not be compatible with

all the processors available. Although the work in Alahmad and Gopalakrishnan (2019) has some connections to the MC task model, they did not explicitly concern different criticality levels of a task. It aims to achieve higher design assurance levels by using adequate monitoring and improving mechanisms. In contrast, the MC scheduling that we propose focuses on providing service guarantee only to the high criticality jobs where computational resources are not adequate. Kato and Ishikawa (2009) introduced gang task scheduling based on global EDF. Dong and Liu (2017) proposed a schedulability analysis based on lag-based reasoning. Few other related works, such as Goossens and Bertin (2010), provided schedulability tests for fixed task-priority scheduling of real-time periodic gang tasks.

Although a good number of works studied MC scheduling and parallel tasks scheduling individually, very few works studied the scheduling of MC parallel tasks (Liu et al. 2014; Baruah 2016; Li et al. 2017; Rambo and Ernst 2017). Rambo and Ernst (2017) proposed a replica-aware co-scheduling approach (that is a combination of strict priority preemptive (SPP) policy and gang scheduling policy) for mixed-critical systems. Baruah (2016) and Li et al. (2017) proposed the MC scheduling of DAG models, while (Liu et al. 2014) proposed the MC scheduling of the synchronous task model. Unlike these works, we consider the gang task model, where a task cannot execute if the number of available cores is less than its degree of parallelism. This constraint makes the scheduling problem highly challenging.

7 Conclusion

Parallel computing with real-time constraints is gaining popularity due to its broad applicability and system efficiency (Kato and Ishikawa 2009; Andersson and de Niz 2012; Bonifaci et al. 2013; Li et al. 2014; Goossens and Richard 2016; Dong and Liu 2017; Guo et al. 2017; Bhuiyan et al. 2018; Guo et al. 2019). WCET measurements are pessimistic due to increased uncertainty. So, there is an emerging need to introduce MC into parallel computation models and system designs. We leverage two existing algorithms (EDF-VD and GEDF) to schedule MC gang tasks efficiently in this work. We derive the first speedup bound for GEDF schedulability of (non-MC) gang tasks and further derived the bound for GEDF-VD of MC gang tasks.

Future work This work is an initial step of more substantial efforts in bringing richer system modeling and analysis into the emerging need in many applications for parallel computing and MC. In the future, by ensuring the a priori verification, we plan to consider run-time robustness, with moldable or malleable models for MC gang tasks. In this work, we have considered the preemptive scheduling, where a high priority task can interrupt a running task, and offers more flexibility in task scheduling. However, preemptive scheduling suffers from the switching overhead as a task from running state to ready state and vice-versa. In the future, we plan to study the non-preemptive scheduling (which does not suffer from task switching overhead) and provide a detailed comparison between these two directions. We also plan to evaluate our results (in this paper) by implementations on applicable hardware platforms.

Acknowledgements We thank the reviewers for their constructive feedback to improve this paper. This work is partially supported by NSF Grant CNS-1850851.

References

- Alahmad B, Gopalakrishnan S (2019) Isochronous execution models for high-assurance real-time systems. In: HASE. IEEE
- Andersson B, de Niz D (2012) Analyzing global-EDF for multiprocessor scheduling of parallel tasks. In: OPODIS. Springer
- Awan M, Bletsas K, Souto P, Tovar E (2017) Semi-partitioned mixed-criticality scheduling. In: ARCS. Springer
- Baruah S (2016) The federated scheduling of systems of mixed-criticality sporadic DAG tasks. In: RTSS. IEEE
- Baruah S (2018) Mixed-criticality scheduling theory: scope, promise, and limitations. *IEEE Des Test* 35(2):31–37
- Baruah S, Bonifaci V, D'angelo G, Marchetti-Spaccamela A, Van Der Ster S, Stougie L (2011) Mixed-criticality scheduling of sporadic task systems. In: ESA. Springer
- Baruah S, Bonifaci V, D'Angelo G, Li H, Marchetti-Spaccamela A, Van Der Ster S, Stougie L (2012) The preemptive uniprocessor scheduling of mixed-criticality implicit-deadline sporadic task systems. In: ECRTS. IEEE
- Baruah S, Bertogna M, Buttazzo G (2015a) Multiprocessor scheduling for real-time systems. Springer, New York
- Baruah S, Eswaran A, Guo Z (2015b) MC-Fluid: simplified and optimally quantified. In: RTSS. IEEE
- Bhuiyan A, Guo Z, Saifullah A, Guan N, Xiong H (2018) Energy-efficient real-time scheduling of DAG tasks. *ACM Trans Embed Comput Syst* 17(5):84
- Bhuiyan A, Sruti S, Guo Z, Yang K (2019a) Precise scheduling of mixed-criticality tasks by varying processor speed. In: RTNS
- Bhuiyan A, Yang K, Arefin S, Saifullah A, Guan N, Guo Z (2019b) Mixed-criticality multicore scheduling of real-time gang task systems. In: RTSS. IEEE
- Bhuiyan A, Liu D, Khan A, Saifullah A, Guan N, Guo Z (2020a) Energy-efficient parallel real-time scheduling on clustered multi-core. *IEEE Trans Parallel Distrib Syst* 31(9):2097–2111
- Bhuiyan A, Reghenzani F, Fornaciari W, Guo Z (2020b) Optimizing energy in non-preemptive mixed-criticality scheduling by exploiting probabilistic information. *IEEE Trans Comput-Aided Des Integr Circ Syst* 39(11):3906–3917
- Bolado M, Posadas H, Castillo J, Huerta P, Sanchez P, Sánchez C, Fouren H, Blasco F (2004) Platform based on open-source cores for industrial applications. In: Design, automation and test in Europe conference and exhibition. Proceedings, vol. 2, pp. 1014–1019. IEEE
- Bonifaci V, Marchetti-Spaccamela A, Stiller S, Wiese A (2013) Feasibility analysis in the sporadic DAG task model. In: ECRTS. IEEE
- Burns A, Davis R (2013) Mixed criticality systems-a review. Department of Computer Science, University of York, Tech. Rep pp. 1–69
- Burns A, Davis R (2018) A survey of research into mixed criticality systems. *ACM Comput Surv* 50(6):82
- Capodiceci N, Cavicchioli R, Bertogna M, Paramakuru A (2018) Deadline-based scheduling for GPU with preemption support. In: RTSS. IEEE
- Döbel B, Härtig H, Engel M (2012) Operating system support for redundant multithreading. In: Proceedings of the tenth ACM international conference on Embedded software, pp 83–92
- Dong Z, Liu C (2017) Analysis techniques for supporting hard real-time sporadic gang task systems. In: RTSS. IEEE
- Easwaran A (2013) Demand-based scheduling of mixed-criticality sporadic tasks on one processor. In: RTSS. IEEE
- Elliott GA, Ward BC, Anderson JH (2013) GPUSync: A framework for real-time GPU management. In: RTSS. IEEE
- Engel M, Döbel B (2012) The reliable computing base-a paradigm for software-based reliability. *INFORMATIK* 2012

- Ernst R, Di Natale M (2016) Mixed criticality systems—a history of misconceptions? *IEEE Des Test* 33(5):65–74
- Esper A, Nelissen G, Nélis V, Tovar E (2015) How realistic is the mixed-criticality real-time system model? In: *RTNS*. ACM
- Feitelson DG, Rudolph L (1992) Gang scheduling performance benefits for fine-grain synchronization. *J Parallel Distrib Comput* 16(4):306–318
- Gehring EF, Siewiorek DP, Segall Z (1987) *Parallel processing: the Cm* experience*. Digital Press, Bedford
- Goossens J, Bertin V (2010) Gang FTP scheduling of periodic and parallel rigid real-time tasks. arXiv preprint [arXiv:1006.2617](https://arxiv.org/abs/1006.2617)
- Goossens J, Richard P (2016) Optimal scheduling of periodic gang tasks. *Leibniz Trans Embed Syst* 3(1):04–1
- Guo Z, Bhuiyan A, Saifullah A, Guan N, Xiong H (2017) Energy-efficient multi-core scheduling for real-time DAG tasks
- Guo Z, Bhuiyan A, Liu D, Khan A, Saifullah A, Guan N (2019) Energy-efficient real-time scheduling of DAGs on clustered multi-core platforms. In: *RTAS*. IEEE
- Kato S, Ishikawa Y (2009) Gang EDF scheduling of parallel task systems. In: 30th IEEE real-time systems symposium, pp 459–468. IEEE
- Lee J, Phan K, Gu X, Lee J, Easwaran A, Shin I, Lee I (2014) MC-Fluid: fluid model-based mixed-criticality scheduling on multiprocessors. In: *RTSS*. IEEE
- Li H (2013) *Scheduling mixed-criticality real-time systems*. Ph.D. thesis, The University of North Carolina at Chapel Hill
- Li H, Baruah S (2010) An algorithm for scheduling certifiable mixed-criticality sporadic task systems. In: *RTSS*. IEEE
- Li J, Agrawal K, Lu C, Gill C (2013) Analysis of global EDF for parallel tasks. In: *ECRTS*. IEEE
- Li J, Chen JJ, Agrawal K, Lu C, Gill C, Saifullah A (2014) Analysis of federated and global scheduling for parallel real-time tasks. In: *ECRTS*. IEEE
- Li J, Ferry D, Ahuja S, Agrawal K, Gill C, Lu C (2017) Mixed-criticality federated scheduling for parallel real-time tasks. *Real-Time Syst* 53(5):760–811
- Liu G, Lu Y, Wang S, Gu Z (2014) Partitioned multiprocessor scheduling of mixed-criticality parallel jobs. In: *RTCSA*. IEEE
- Nvidia (2017) <http://www.nvidia.com/page/home.html>
- Openacc (2017) <https://www.openacc.org/>
- Ousterhout JK et al (1982) Scheduling techniques for concurrent systems. *ICDCS* 82:22–30
- Rambo EA, Ernst R (2017) Replica-aware co-scheduling for mixed-criticality. In: *ECRTS 2017*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik
- Spuri M, Buttazzo GC (1994) Efficient aperiodic service under earliest deadline scheduling. In: *RTSS*, pp. 2–11
- Stavrinides GL, Karatza HD (2016) Scheduling real-time parallel applications in saas clouds in the presence of transient software failures. In: *SPECTS*. IEEE
- Tobuschat S, Ernst R (2017) Efficient latency guarantees for mixed-criticality networks-on-chip. In: *RTAS*. IEEE
- Trüb R, Giannopoulou G, Tretter A, Thiele L (2017) Implementation of partitioned mixed-criticality scheduling on a multi-core platform. *ACM Trans Embed Comput Syst* 16(5s):122
- Vestal S (2007) Preemptive scheduling of multi-criticality systems with varying degrees of execution time assurance. In: *RTSS*. IEEE
- Xiao S, Feng Wc (2010) Inter-block GPU communication via fast barrier synchronization. In: *IPDPS*. IEEE
- Yang M, Amert T, Yang K, Otterness N, Anderson JH, Smith FD, Wang S (2018) Making OpenVX really “real time”. In: *RTSS*. IEEE

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Ashikahmed Bhuiyan is a Ph.D. candidate in the Department of Electrical and Computer Engineering at the University of Central Florida (UCF), under the supervision of Zhishan Guo and Abusayeed Saifullah (from the Wayne state University). He is a member of the Real-Time & Intelligent Systems Lab at UCF. He received his Bachelor of Science degree in Computer Science and Engineering from Bangladesh University of Engineering and Technology (BUET), Bangladesh, in 2013. His research focuses on improving energy efficiency in real-time embedded systems, parallel computing, and mixed-criticality scheduling. He has received the Best Student Paper Award at the 40th IEEE Real-Time Systems Symposium (RTSS 2019).



Kecheng Yang is an Assistant Professor in the Department of Computer Science at Texas State University. He received his Ph.D. and M.S. degrees in Computer Science from the University of North Carolina at Chapel Hill in 2018 and 2015, respectively, both with Prof. James H. Anderson. Before that, he received his B.E. degree in Computer Science and Technology from Hunan University in 2013. His research interests include real-time and cyber-physical systems, scheduling theory and resource allocation algorithms, and heterogeneous multiprocessor platforms.



Samsil Arefin is a software engineer in Microsoft New England Research and Development Center. He received his B.Sc. degree in Computer Science and Engineering from Bangladesh University of Engineering and Technology (BUET). He received his M.S. in Computer science from Missouri University of Science & Technology under the supervision of Dr. Zhishan Guo. His research focuses on realtime mixed-criticality task scheduling and graceful degradation.



Abusayeed Saifullah is an assistant professor of the Computer Science Department at Wayne State University. He received PhD in Computer Science and Engineering with Turner Dissertation Award from Washington University in St Louis in 2014. His research primarily concerns Internet-of-Things, cyber-physical systems, real-time systems, embedded systems, and lowpower wide-area networks. He received 8 Best Paper Awards and nominations in highly competitive conferences including EWSN (2021 nomination), ACM SenSys (2016 nomination), IEEE RTSS (2019, 2014, 2011), IEEE ICI (2018), and IEEE RTAS (2012 nomination). He also received multiple young investigator awards including the CAREER award (2019) and the CRII award (2016) of the National Science Foundation (NSF). He is an editor of Elsevier Pervasive and Mobile Computing journal.



Nan Guan is currently an Associate Professor at City University of Hong Kong. He got his PhD in Uppsala University, Sweden. His research interests include the design and analysis of real-time systems, embedded systems, cyberphysical systems and Internet-of-Things (IoT) systems. He received the EDAA Outstanding Dissertation Award in 2014, the CCF Outstanding Dissertation Award in 2013, the Best Paper Award of IEEE RTSS in 2009, the Best Paper Award of DATE in 2013, the Best Paper Award of ACM e-Energy 2018 and the Best Paper Award of IEEE ISORC 2019, the Outstanding Paper Award of RTSS 2019 and the Best Paper Award of EMSOFT 2020.



Zhishan Guo is an assistant professor in the Department of Electrical and Computer Engineering at The University of Central Florida. He received his B.E. degree (with honor) in computer science and technology from Tsinghua University, Beijing, China, in 2009, his M.Phil. degree in mechanical and automation engineering from the Chinese University of Hong Kong in 2011, and his Ph.D. degree in computer science from the University of North Carolina at Chapel Hill in 2016. His research and teaching interests include real-time scheduling, cyber-physical systems, and neural networks and their applications.

Authors and Affiliations

Ashikahmed Bhuiyan¹  · **Kecheng Yang**² · **Samsil Arefin**³ ·
Abusayeed Saifullah⁴ · **Nan Guan**⁵ · **Zhishan Guo**¹

Ashikahmed Bhuiyan
ashik@knights.ucf.edu

Kecheng Yang
yangk@txstate.edu

Samsil Arefin
Samsil.Arefin@microsoft.com

Abusayeed Saifullah
saifullah@wayne.edu

Nan Guan
nanguan@cityu.edu.hk

¹ University of Central Florida, Orlando, USA

² Texas State University, San Marcos, USA

³ Microsoft New England Research and Development Center, Cambridge, USA

⁴ Wayne State University, Detroit, USA

⁵ City University of Hong Kong, Kowloon, Hong Kong