1 Introduction

In many real-time and embedded systems, tasks may be suspended by the operating system when accessing external devices such as disks, graphical processing units (GPUs), or synchronizing with other tasks in distributed or multicore systems. This behavior is often known as *self-suspension.* Self-suspensions are even more pervasive in many emerging embedded cyber-physical systems in which the computation components frequently interact with external and physical devices [14,15]. Typically, the resulting suspension delays range from a few microseconds (e.g., a write operation on a flash drive [14]) to a few hundreds of milliseconds (e.g., offloading computation to GPUs [15,23]).

The self-suspending task model [**?**] is an useful and widely studied model that can accurately convey the characteristics of many real-time embedded systems that are often seen in practice. The self-suspending task model can be used to represent systems where tasks may experience suspension delays when being blocked to access external devices and shared resources [**?**,**?**,**?**,**?**]. For example, suspension delays introduced by accessing devices such as GPUs could range from a few milliseconds to several seconds [**?**].

Applications of self-suspending task models and the importance: computation offloading, I/O intensive applications, multicore synchronisations, task-graph scheduling. Giorgio suggested us to point out the wide applications of self-suspending task models.

2 Self-Suspending Sporadic Real-Time Task Models

Self-suspending tasks can be classified into two models: *dynamic* self-suspension and *segmented* (or *multi-segment*) self-suspension models. The dynamic self-suspension sporadic task model characterizes each task τ_i as a 4-tuple (C_i, S_i, T_i, D_i): T_i denotes the minimum inter-arrival time (or period) of τ_i, D_i is the relative deadline, C_i denotes the upper bound on total execution time of each job of τ_i, and S_i denotes the upper bound on total suspension time of each job of τ_i. In addition to the above 4-tuple, the segmented sporadic task model further characterizes the computation segments and suspension intervals as an array $(C_i^1, S_i^1, C_i^2, S_i^2, ..., S_i^{m_i-1}, C_i^{m_i})$, composed of m_i computation segments separated by $m_i - 1$ suspension intervals.

From the system designer's perspective, the dynamic self-suspension model provides an easy way to specify self-suspending systems without considering the juncture of I/O access, computation offloading, or synchronization. However, from the analysis perspective, such a dynamic model leads to quite pessimistic results in terms of schedulability since the location of suspensions within a job is oblivious. Therefore, if the suspending patterns are well-defined and characterized with known suspending intervals, the multi-segment self-suspension task model is more appropriate.

definition of static-, dynamic-priority scheduling, schedulability, response time, worst-case response time, etc.

2.1 Examples of Dynamic Self-Suspension Model

different program paths
self-suspension due to synchronizations
etc.

2.2 Examples of Segmented Self-Suspension Model

static execution patterns
multiprocessor synchronization with critical sections
etc.

3 General Design and Analysis Strategies in Uniprocessor Platforms

This section reviews the existing solutions for scheduling and analyzing the schedulability of self-suspending task models. We will first explain the commonly adopted strategies in those solutions. The strategies are generally correct, but the analysis has to be done carefully. In the next section, we will explain some of misconceptions used in the literature by giving concrete reasons and some counterexamples to explain why such misconceptions may lead to over-optimistic analysis. At the end of this section, we will provide the rule of thumb for analyzing self-suspending task systems.

To demonstrate how the scheduling algorithms and the schedulability tests work in existing approaches, we will mainly use the following tasks in Table 1 and Table 2. For demonstrating the worst-case response time analysis, we leave some relative deadline with "?" and period ∞. Specifically, we will use task set $\mathbf{T}_1 = \{\tau_1, \tau_2, \tau_3\}$, $\mathbf{T}_2 = \{\tau_1, \tau_2, \tau_3, \tau_4\}$, $\mathbf{T}_3 = \{\tau_\alpha, \tau_\beta, \tau_\gamma\}$ in our examples. Unless specified, we will implicitly assume that these three example task sets are scheduled under Rate-Monotonic scheduling.

	C_i	S_i	D_i	T_i
τ_α	1	0	2	2
τ_β	5	5	20	20
τ_γ	1	0	?	∞

Table 1. Examples for dynamic self-suspending tasks

For self-suspending sporadic task systems, while executing, a job may suspend itself or even must suspend itself in the segmented self-suspension model. While a job suspends, the scheduler removes the job from the ready queue. Such suspensions should be well characterized and the resulting workload interference should be well quantified to analyze the schedulability of the task systems.

an example

	(C_i^1, S_i^2, C_i^2)	D_i	T_i
τ_1	(2, 0, 0)	5	5
τ_2	(2, 0, 0)	10	10
τ_3	(1, 5, 1)	15	15
τ_4	(3, 0, 0)	?	∞

Table 2. Examples for dynamic segmented-suspending tasks

There are some common strategies to characterize and quantify the impact due to self-suspensions:

- **Convert All Self-Suspension into Computation:** This is the simplest and the most pessimistic strategy. It basically converts all self-suspending time into computation time. That is, we can consider that the execution time of task τ_i is always $C_i + S_i$. After the conversion, we only have sporadic real-time tasks. Therefore, all the existing results for sporadic task systems can be adopted. The proof can be done with the following simple interpretation: The suspension of a job may make the processor idle. If two jobs suspend at the same time and the processor idles in a certain time interval in the actual schedule, it can be imagined that one of these two jobs have shorter execution time (than its worst-case execution time $C_i + S_i$). Such earlier completion does not affect the schedulability analysis. Therefore, putting $C_i + S_i$ as the worst-case execution time for every task τ_i is a very safe analysis for both dynamic- and static-scheduling policies. Such an approach has been widely used as the baseline of more accurate analyses in the literature.
  With this schedulability test, it is easy to see that none of the three example task sets $\mathbf{T}_1$, $\mathbf{T}_2$, $\mathbf{T}_3$ cannot be classified as feasible since $\frac{1}{2} + \frac{5+5}{20} + \frac{1}{D_\gamma} > 1$ and $\frac{2}{5} + \frac{2}{10} + \frac{1+5+1}{15} > 1$.
- **Convert Higher-Priority Tasks into Sporadic Tasks:** In static-priority scheduling, we can convert the higher-priority self-suspending tasks into equivalent sporadic real-time tasks: When we analyze the schedulability of a task τ_k, we can convert the higher-priority self-suspending tasks into sporadic tasks by treating the suspension as computation. That is, a higher-priority task τ_i (higher than task τ_k) has now worst-case execution time $C_i + S_i$. This simplifies the analysis. After converting, we only have one self-suspending task left as the lowest-priority task in the system. Such a conversion is useful for analyzing segmented self-suspending task model. However, such a conversion is not very useful for analyzing dynamic self-suspending task models, since we have to consider the worst case that the self-suspension of task τ_k makes the processor idle. Therefore, we also have to convert τ_k's self-suspension into computation. This results in identical analysis by converting self-suspension into computation for all the tasks.
  With the conversion, the fundamental problem is to analyze the worst-case response time of a self-suspending task τ_k as the lowest-priority task in the task system, when all the other higher priority tasks are ordinary sporadic real-time tasks. One simple strategy is to analyze the worst-case response

time R_k^j for each computation segment C_k^j. The schedulability test of task τ_k then is to simply verify whether $R_k^{m_k} + \sum_{j=1}^{m_k-1} R_k^j + S_k^j \leq D_k \leq T_k$. Let's use task set $\mathbf{T}_1$ as an example. The worst-case response times of $C_3^1 = 1$ and $C_3^2 = 1$ in $\mathbf{T}_1$ are both clearly 5 by using standard the time demand analysis (TDA). Therefore, we know that the worst-case response time of task τ_3 in $\mathbf{T}_1$ is at most 15.

The above test can be pretty pessimistic especially when the suspending time is short. Imagine that we change S_3^1 from 5 to 1. The above analysis still considers that both computation segments suffer from the worst-case interference from the two higher-priority tasks and returns 11 as the (upper bound of the) worst-case response time. For this new configuration, if we greedily convert the suspension into computation and use TDA analysis, we can conclude that the worst-case response time is at most 9.

Therefore, it can be more precise if the higher-priority interference is analyzed more precisely. But, it has to be done carefully. The problem with only one self-suspending task as the lowest-priority task has been specifically studied in [20,9]. Unfortunately, the analysis in [20] is flawed. We will explain the reasons in Section 4.1.

– **Quantify Additional Interference due to Self-Suspensions:** Suspension may result in more workload from higher-priority jobs to interfere with a lower-priority job. This strategy is to convert the suspension time of a job of task τ_k under analysis into computation. Suppose that this job under analysis arrives at time t_k. The other higher-priority jobs except the job under analysis are considered to (possibly) have self-suspensions. This is the completely opposite strategy to the previous strategy. Since a higher-priority self-suspending job may suspend itself before t_k and resume after t_k, the self-suspending behaviour of a task τ_i can be considered to bring *at most* one *carry-in* job to be *partially* executed after t_k if $D_i \leq T_i$. As we have converted task τ_k's self-suspension time into computation, the finishing time of the job of task τ_k is the earliest moment after t_k such that the processor idles.

  • In the *dynamic self-suspending task model*, the above analysis implies that the higher-priority jobs arrived after time t_k *should not* suspend themselves to create the maximum interference. Therefore, suppose that the first arrival time of task τ_i after t_k is t_i, i.e., $t_i \geq t_k$. Then, the demand of task τ_i released at time $t \geq t_i$ is $\left\lceil \frac{t-t_i}{T_i} \right\rceil C_i$. So, we just have to account the demand of the carry-in job of task τ_i executed between t_k and t_i. The workload of the carry-in job can be up to C_i, but can also be characterized in a more precise manner. The approaches in this category are presented in [13,21] by greedily counting C_i in the carry-in job. Jane W.S. Liu in her textbook [22, Page 164-165] presents an approach to quantify the higher-priority tasks by setting up the *blocking time* induced by self-suspensions. In her analysis, a job of task τ_k can suffer from the *extra delay* due to self-suspending behavior as a factor of blocking time, denoted as B_k, as follows: (1) The blocking time contributed from task

τ_k is S_k. (2) A higher-priority task τ_i can only block the execution of task τ_k by at most $b_i = min(C_i, S_i)$ time units. In the textbook [22], the blocking time $B_k = S_k + \sum_{i=1}^{k-1} b_i$ is then used to perform utilization-based analysis for rate-monotonic scheduling. However, there was no proof in the textbook. Fortunately, the recent report from Chen [XXX] has provided a proof to support the correctness of the above method in [22].
Let's use task set $\mathbf{T}_3$ to illustrate the above analysis in [22, Page 164-165]. In this case, b_β is 5. Therefore, $B_\gamma = 5$. So, the worst-case response time of task τ_γ is upper bounded by the minimum t with $t = B_\gamma + C_\gamma + \left\lceil \frac{t}{T_\alpha} \right\rceil C_\alpha + \left\lceil \frac{t}{T_\beta} \right\rceil C_\beta = 6 + \left\lceil \frac{t}{2} \right\rceil 1 + \left\lceil \frac{t}{20} \right\rceil 5$. The above equality holds when $t = 32$. Therefore, the worst-case response time of task τ_γ in $\mathbf{T}_3$ is upper bounded by 32.

Another way to quantify the impact is to model the impact of the carry-in job by using the concept of *jitter*. If the jitter of task τ_i to model self-suspension is J_i, then, the demand of task τ_i released from $t_i - T_i$ up to time $t + t_k$ (i.e., the demand that can be executed from t_k to $t_k + t$) is $\left\lceil \frac{t+J_i}{T_i} \right\rceil C_i$. A safe way it to set J_i to T_i, which can be imagined as a pessimistic analysis by assuming that the carry-in job of task τ_i has execution time C_i and the release time t_i is t_k. A more precise way to quantify the jitter is to use the worst-case response time of a higher-priority task τ_i. Therefore, we can set the jitter J_i of task τ_i to $T_i - C_i$ [13,27] or $R_i - C_i$ [13], where R_i is the worst-case response time of a higher-priority task τ_i.
Let's use task set $\mathbf{T}_3$ to illustrate the above analysis in [13]. In this case, J_β is $20 - 5 = 15$. So, the worst-case response time of task τ_γ is upper bounded by the minimum t with $t = C_\gamma + \left\lceil \frac{t}{T_\alpha} \right\rceil C_\alpha + \left\lceil \frac{t+15}{T_\beta} \right\rceil C_\beta = 1 + \left\lceil \frac{t}{2} \right\rceil 1 + \left\lceil \frac{t+15}{20} \right\rceil 5$. The above equality holds when $t = 22$. Therefore, the worst-case response time of task τ_γ in $\mathbf{T}_3$ is upper bounded by 22.
There have been some flawed analyses in the literature [2,3,17] which quantify the jitter of task τ_i by setting J_i to S_i. We will explain later in Section 4.3 why setting J_i to S_i is in general too optimistic.

- In the *segmented self-suspending task model*, we can simply ignore the segmentation structure of computation segments and suspension intervals and directly apply all the strategies for dynamic self-suspending task models. However, the analysis will become pessimistic. This is due to the fact that the segmented-suspensions are not completely dynamic. The static suspension patterns result in also certain (more predictable) suspension patterns. However, characterizing the worst-case suspending patterns of the higher priority tasks to quantify the additional interference under segmented self-suspending task model is not easy. Similarly, one possibility is to characterize the worst-case interference in the carry-in job of a higher-priority task τ_i by analyzing its self-suspending pattern, as presented in [12]. Another possibility is to quantify the interference by modeling it with a jitter term, as presented in [4]. We will explain later in Section 4.2 why the quantification of the interference in [4] is incorrect. **Michael's paper in RTSS1998**.

Let's use task set $\mathbf{T}_2$ to illustrate how the schedulability tests work. jj: leave to Kevin and Michael. : endjj

– **Handle Self-Suspension Segments of the Task under Analysis:** Greedily converting the suspension time of a job of task τ_k under analysis into computation can also become very pessimistic if S_k is much larger than C_k. However, the decision to convert a task τ_k has to be done carefully. Now, we can consider a simple example to analyze the worst-case response time of task $\tau_k = ((C_k^1, S_k^1, C_k^2), T_k, D_k)$. We can have two options:
  - Convert S_k^2 into computation, and then apply the above analysis by considering that task τ_k has execution time $C_k^1 + S_k^1 + C_k^2$. We simply have to verify whether the worst-case response time is no more than D_k.
  - Treat each of the computation segments of task τ_k individually by applying the worst-case higher-priority interference, regardless of its previous computation segments. We need to verify if the suspension time S_k plus sum of the worst-case response time of all the computation segments of task τ_k is no more than D_k.

  The benefit of the former approach is due to that it only pessimistically counts the additional higher-priority interference once. However, it also suffers from the pessimism by converting S_k^1 into computation. The benefit of the latter approach is due to the fact that the suspension time is not over-counted as computation. However, it also over-counts the carry-in workload since every computation segment may have to pessimistically count the worst-case workload of the carry-in jobs. Both of these two approaches are adopted in the literature [9,12,4]. They can be both applied and the better result is returned.

  The example in Convert Higher-Priority Tasks into Sporadic Tasks when S_3^1 is 1 has demonstrated the difference of the above two difference cases.

– **Enforce Periodic Behaviour by Release Time Enforcement:** Self-suspension can cause substantial schedulability degradation. To leviate the impact on additional interference due to self-suspension, one possibility is to enforce the periodic behaviour by enforcing the release time of the computation segments. There are two categories of such enforcement.
  - *Use resource reservation servers*: Rajkumar [27] proposes a *period enforcer* algorithm to handle the impact of uncertain releases (like self-suspensions). jj: rewrite this... : endjj (One can imagine that a sporadic server [29] with capacity C_k and replenishment period T_k is the reservation server to run task τ_k, by handling the self-suspension.) The period enforcer algorithm was shown the have a good property: "A deferrable task that is schedulable under its worst-case conditions is also schedulable under the period enforcer algorithm." in Theorem 3.5 in [27]. jj: leave this to Raj. : endjj
  - *Set a constant offset to constrain the release time of a computation segment*: Suppose that the offset for the j-th computation segment of task

τ_i is ϕ_i^j. This means that the j-th computation segment of task τ_i is released only at time $r_i + \phi_i^j$, in which r_i is the arrival time of a job of task τ_i. With the enforcement, each computation segment can be represented by a sporadic task with a period T_i, a WCET C_i^j, and a relative deadline $\phi_{i,j+1} - \phi_i^j - S_i^j$. (Here, ϕ_{i,m_i+1} is set to D_i.) Such approaches have been presented in [18,20,7]. The method in [7] is a simple greedy solution for implicit-deadline self-suspending task systems with at most one self-suspension interval per task. It assigns the phase ϕ_i^2 always to $\frac{T_i+S_i^1}{2}$ and the relative deadline of the first computation segment of task τ_i to $\frac{T_i-S_i^1}{2}$. This is the first method in the literature with *speedup factor* guarantees by using the revised relative deadline for earliest-deadline-first scheduling.

The method in [18] assigns each computation segment a static-priority level and a phase. Unfortunately, in [18], the schedulability test is not correct, and the proposed mixed-integer linear programming is unsafe for worst-case response time guarantees. The method in [20] is XXX (left to Geoffrey...)

– **Special Cases with Good Observations:**

4 Misconceptions in Some Existing Results

4.1 Incorrect Assumptions in Critical Instant Theorem with Synchronous Releases

4.2 Incorrect Quantifications of Additional Interferences due to Carry-In Jobs

4.3 Incorrect Quantifications of Jitter

4.4 Incorrect Periodic Execution Enforcement

5 Self-Suspending Tasks in Multiprocessor Synchronizations

In this section, we consider multiprocessors subject to *partitioned fixed-priority (P-FP)* scheduling, and review the general analysis strategies for tasks that synchronize access to shared resources (*e.g.*, shared I/O devices, communication buffers, or scheduler locks) with suspension-based locks (*e.g.*, binary semaphores). Unfortunately, some of the misconceptions surrounding the analysis of self-suspensions on uniprocessors also spread to the analysis of partitioned multiprocessor real-time locking protocols. In particular, as we show with a counterexample, the analysis framework to account for the additional interference due to *remote blocking* first introduced in [19], and reused in several other works [33,5,30,16,10,6,31], is flawed. Finally, a straightforward solution for these problems are discussed.

5.1 Existing analysis strategies

P-FP scheduling is a widespread choice in practice due to the wide support in industrial standards such as AUTOSAR, and in many RTOSs like VxWorks, RTEMS, ThreadX, *etc.*Under P-FP scheduling, each task has a fixed base priority and is statically assigned to a specific processor, and the tasks on each processors are scheduled as in uniprocessors.

Under partitioned scheduling, a resource accessed by tasks from different processors is called a *global resource*, otherwise it is called a *local resource*. When a job requests a global resource, it may incur *remote blocking* if the global resource is held by a job on another processor. Also, a job may incur *local blocking* if it is prevented from being scheduled by a resource-holding job of a lower-priority task on its local processor.

Under suspension-based protocols, such as the *multiprocessor priority ceiling protocol (MPCP)* [26], tasks that are denied to access shared resources are suspended. From the perspective of local schedule on each processor, remote blocking, caused by external events (*i.e.*, resource contention due to tasks on the other processors), pushes the execution of higher-priority tasks to a later time point regardless of the schedule on the local processor (*i.e.*, even if the local processor is idle), thus may cause additional interference on lower-priority tasks. To this end, remote blocking is considered as self-suspension in analysis. Whereas, local blocking takes place only if a local lower-priority task is scheduled (*i.e.*, the local processor is busy). Consequently, local blocking is accounted for as regular blocking as in uniprocessors, but not as self-suspension.

In analysis, a safe yet pessimistic strategy is to convert remote blocking into computation. Accordingly, the remote blocking incurred by each higher-priority task is counted as part of interference. block-2007 first used this strategy for partitioned *earliest deadline first (EDF)* scheduling; lakshmanan-2009 also adopted this approach in their analysis of "virtual spinning," where when a task is suspended due to remote blocking other tasks are allowed to execute unless they try to access global resources. An alternative is to bound the effects of deferred execution due to remote blocking. Recently, lakshmanan-2009 proposed the following response-time analysis framework that takes into account the amount of remote blocking to bound the worst case interference.

$$R_k^{n+1} = C_k^\star + \sum_{\tau_i \in hp(k) \cap P(\tau_k)} \left\lceil \frac{R_k^n + B_i^r}{T_i} \right\rceil \cdot C_i + s_k \sum_{\tau_l \in lp(k) \cap P(\tau_k)} \max_{1 \leq j < s_l} C'_{l,j}. \quad (1)$$

where $C_k^\star = C_k + B_k^r$, $R_k^0 = C_k^\star$, and τ_k is considered schedulable if $R_k^{n+1} = R_k^n < D_k$.

In Eq. 1, B_k^r is an upper bound on the maximum remote blocking that a job of τ_k incurs. $hp(k)$ and $lp(k)$ denote the tasks with higher and lower priority than τ_k, respectively. $P(\tau_k)$ denotes the tasks that are assigned on the same processor as τ_k. s_k is the maximum number of critical sections of τ_k. $C'_{l,j}$ is an upper bound on the execution time of the jth critical section of τ_l. The additional interference on τ_k due to the remote blocking incurred by each higher-priority task is captured in the analysis according to the second term in Eq. 1.

Under this analysis framework, $\left\lceil \frac{R_k+B_i^r}{T_i} \right\rceil \cdot C_i$ is used as an upper bound on the total interference of τ_i on τ_k. This analysis approach was reused in [30,16,6,31] to analyze (several variants of) the MPCP. The response-time analysis framework (*i.e.*, Eq. 1) was also reused in [33,5,10] to compare the schedulability performances between different locking protocols under P-FP scheduling.

Unfortunately, the analysis approach based on Eq. 1 fails to guarantee a safe response time bound in certain corner cases, as can be demonstrated with the following counterexample.

5.2 A counterexample

We show the existence of a schedule in which a task that is considered schedulable according to the analysis in [19] is in fact unschedulable.

τ_k	C_k	T_k $(= D_k)$	$N_{k,1}$	$L_{k,1}$	$N_{k,2}$	$L_{k,2}$
τ_1	2	6	0	0	0	0
τ_2	$4+2\epsilon$	13	1	ϵ	0	0
τ_3	ϵ	14	0	0	1	ϵ
τ_4	6	14	1	$4-\epsilon$	1	ϵ

Table 3. Task parameters

Consider four implicit deadline sporadic tasks $\tau_1, \tau_2, \tau_3, \tau_4$ (with parameters listed in Table 3, where $N_{k,1}$ $(N_{k,2})$ denotes the maximum number of requests that a job of τ_k can issue to resource ℓ_1 (ℓ_2), and $L_{k,1}$ $(L_{k,2})$ denotes the corresponding maximum critical section length), ordered by decreasing order of priority, that are scheduled on two processors using P-FP scheduling. τ_1, τ_2 and τ_3 are assigned to processor 1, while T_4 is assigned to processor 2. Jobs of τ_2 and τ_4 each once access a shared resource ℓ_1 $(N_{2,1} = 1$ and $N_{4,1} = 1)$. Jobs of τ_4 each once access a shared resource ℓ_2 $(N_{4,2} = 1)$. Each job of τ_2 uses ℓ_1 for a duration of at most $L_{2,1} = \varepsilon < 1$ time units (an arbitrarily small quantity), and each job of τ_4 uses ℓ_1 and ℓ_2 for at most $L_{4,1} = 4 - \varepsilon$ and $L_{4,1} = \varepsilon$ time units respectively.

Consider the response-time of τ_3. Since τ_3 is the lowest-priority task on its processor, it does not incur any local blocking ($i.e.$,$s_3 \sum_{\tau_l \in lp(3) \cap P(\tau_3)} \max_{1 \le j < s_l} C'_{l,j} =$ 0). While each time τ_3 requests ℓ_2, it may be delayed by τ_4 for at most ϵ. Thus, the maximum remote blocking of τ_3 is bounded by $B_3^r = \epsilon$ [1]. With regard to the remote blocking incurred by each higher-priority task, we have $B_1^r = 0$ because

[1] In general, the upper bound on blocking of course depends on the specific locking protocol in use, but in this example, by construction, the stated bound holds under any reasonable locking protocol. Recent surveys of multiprocessor semaphore protocols may be found in [5,32].

τ_1 does not request any global resource. Further, each time when a job of T_2 requests ℓ_1, it may be delayed for a duration of at most $4-\varepsilon$, thus $B_2^r = 4-\varepsilon$. Therefore, according to Eq. 1, we have

$$R_3^0 = \varepsilon + \varepsilon = 2\varepsilon,$$
$$R_3^1 = 2\varepsilon + 0 + \left\lceil \frac{2\varepsilon + 0}{6} \right\rceil \cdot 2 + \left\lceil \frac{2\varepsilon + 4 - \varepsilon}{13} \right\rceil \cdot (4 + 2\varepsilon) = 2\varepsilon + 1 \cdot 2 + 1 \cdot (4 + 2\varepsilon) = 6 + 4\varepsilon,$$
$$R_3^2 = 2\varepsilon + 0 + \left\lceil \frac{6 + 4\varepsilon + 0}{6} \right\rceil \cdot 2 + \left\lceil \frac{6 + 4\varepsilon + 4 - \varepsilon}{13} \right\rceil \cdot (4 + 2\varepsilon) = 2\varepsilon + 2 \cdot 2 + 1 \cdot (4 + 2\varepsilon) = 8 + 4\varepsilon,$$
$$R_3^3 = 2\varepsilon + 0 + \left\lceil \frac{8 + 4\varepsilon + 0}{6} \right\rceil \cdot 2 + \left\lceil \frac{8 + 4\varepsilon + 4 - \varepsilon}{13} \right\rceil \cdot (4 + 2\varepsilon) = 2\varepsilon + 2 \cdot 2 + 1 \cdot (4 + 2\varepsilon) = 8 + 4\varepsilon.$$

As a result, $R_3 = 8 + 4\varepsilon < 14 = D_3$, and τ_3 is considered to be schedulable according to the analysis in [19]. However, there exists a schedule, shown in Fig. **??**, where τ_3 actually misses a deadline at time 20, which implies that the analysis framework in [19] is too optimistic (in certain cases).

5.3 Incorrect Time Request Analysis With Global Resource Sharing

Besides the over-optimistic problem found in lakshmanan-2009, a straightforward adoption of the traditional time request analysis with global resource sharing is also not safe, as we show next.

In [24], an *interface-based analysis* based on the time request analysis was derived for real-time open systems, where each processor contains a task set and the tasks on different processors may share global resources. Intuitively, the interface abstracts the requirements of global resource sharing on each processor that should be satisfied to guarantee the schedulability of the system in the processor. In particular, the maximum blocking time that a task τ_k can tolerate without missing its deadline, denoted by $mtbt_k$, is evaluated as following.

Starting from the schedulability condition, τ_k is considered schedulable in a single processor if

$$\exists t \in (0, D_k] : rbf_{FP}(k,t) \leq t, \tag{2}$$

where $rbf_{FP}(k,t)$ is the *request bound function* of τ_k, which is computed by

$$rbf_{FP} = C_k + B_k + \sum_{\tau_i \in hp(k)} \left\lceil \frac{t}{T_i} \right\rceil \cdot C_i. \tag{3}$$

In Eq. 3, B_k denotes the total blocking time of τ_k. By substituting B_k and $mtbt_k$, $mtbt_k$ is then computed by

$$mtbt_k = \max_{0<t\leq D_k} \left(t - (C_k + \sum_{\tau_i \in hp(k)} \left\lceil \frac{t}{T_i} \right\rceil \cdot C_i) \right). \tag{4}$$

However, from to the analysis in section 4.2, we can already infer that the analysis used in Eq. 2 and Eq. 3, which ignores the effects of remote blocking on interference and quantifies the total interference from a task τ_i on τ_k over a duration of length t as $\left\lceil \frac{t}{T_i} \right\rceil \cdot C_i$, is over optimistic.

Consider τ_3 in the previous example (with parameters listed in Table 3). According to Eq. 4, $mtbt_3$ is $12 - (\epsilon + \lceil 12/6 \rceil \cdot 2 + \lceil 12/13 \rceil \cdot (4 + 2\epsilon)) = 4 - 3\epsilon$ (when $t = 12$), which implies that τ_3 can tolerate a maximum blocking time of at least $4 - 3\epsilon$ without missing its deadline. However, this is not true since τ_3 is unschedulable even without incurring any blocking, as shown in Fig. **??**.

5.4 A Safe Response Time Bound

In Eq. 1, the remote blocking of each higher-priority task (*i.e.*, B_i^r) is counted in a similar way as release jitter. However, it is not sufficient to count the duration of remote blocking as release jitter (already explained in section 3.2.3). A straightforward fix is thus to replace B_i^r, in the ceiling function (*i.e.*, the second term in Eq. 1), with a larger value such as D_i (as proved/discussed in section 3.3) or $R_i - C_i$ (as proved / discussed in section 3.3). Similarly, replacing $\sum_{\tau_i \in hp(k)} \lceil t/T_i \rceil \cdot C_i$ in Eq. 3 and Eq. 4 with $\sum_{\tau_i \in hp(k)} \lceil (t + D_i)/T_i \rceil \cdot C_i$ or $\sum_{\tau_i \in hp(k)} \lceil (t + R_i - C_i)/T_i \rceil \cdot C_i$ may fix the over-optimistic problem in [24].

Further, since most papers reviewed in section 4.1 merely reused the over-optimistic analysis framework in [19], the stated fix may be used to correct the response-time tests in these papers.

6 Hardness Review of Self-Suspending Task Models

This section reviews the hardness for designing scheduling algorithms and schedulability analysis of self-suspending task systems.

6.1 Hardness for Scheduling Segmented Self-Suspending Tasks

Verifying the existence of a feasible schedule for segmented self-suspending task systems is proved to be $\mathcal{NP}$-hard in the strong sense in [28]. It is also shown that EDF and RM do not have any speedup factor bound in in [28] and [7], respectively.

The only results with speedup factor analysis for fixed-priority scheduling and dynamic priority scheduling can be found in [7] and [11]. The analysis with speedup factor 3 in [7] can be used for systems with at most one self-suspension interval per task in dynamic priority scheduling. The analysis with a bounded speedup factor in [11] can be used for fixed-priority and dynamic-priority systems with any number of self-suspension intervals per task. However, the speedup factor in [11] depends on, and grows quadratically with respect to the number of self-suspension intervals. Therefore, it can only be *practically* used when there are only a few number of suspension intervals per task. The scheduling policy used in [11] is *laxity-monotonic* (LM) scheduling, which assigns the highest priority to the task with the least laxity, that is, $D_i - S_i$.

With respect to this scheduling problem, there was no theoretical lower bound (with respect to the speedup factors) of this scheduling problem.

The above analysis also implies that the priority assignment in fixed-priority scheduling should be carefully designed. Traditional approaches like RM or EDF do not work very well. LM may work for a few self-suspending intervals, but how to perform the optimal priority assignment is an open problem.

6.2 Hardness for Scheduling Dynamic Self-Suspending Tasks

The complexity class for verifying the existence of a feasible schedule for dynamic self-suspending task systems is unknown in the literature. The proof in [28] cannot be applied to this case. It is proved in [13] that the speed-up factor for RM, DM, and LM scheduling is ∞. Here, we repeat the example in [13]. Consider the following implicit-deadline task set with one self-suspending task and one sporadic task:

- $C_1 = 1 - 2\epsilon$, $S_1 = 0$, $T_1 = 1$
- $C_2 = \epsilon$, $S_2 = T - 1 - \epsilon$, $T_2 = T$

where T is any natural number larger than 1 and ϵ can be arbitrary small.

It is clear that this task set is schedulable if we assign higher priority to task τ_2. Under either RM, DM, and LM scheduling, task τ_1 has higher priority than task τ_2. It was proved in [13] that this example has a speed-up factor ∞ when ϵ is close to 0.

There is no upper bound of this problem in the most general case. The analysis in [13] for a speedup factor 2 uses a trick to compare the speedup factor with respect to the *optimal fixed-priority schedule* instead of the *optimal schedule*. There is no proof or evident to show that this factor 2 is also the factor when the reference is the *optimal schedule*.

With respect to this problem, there was no theoretical lower bound (with respect to the speedup factors) of this scheduling problem.

The above analysis also implies that the priority assignment in fixed-priority scheduling should be carefully designed. Traditional approaches like RM or EDF do not work very well. LM also does not work well. The priority assignment used in [13] is based on the optimal-priority algorithm (OPA) from Audsley [1] with an OPA-compatible schedulability analysis. However, since the schedulability test used in [13] is not exact, the priority assignment is also not the optimal solution. Finding the optimal priority assignment here is also an open problem.

6.3 Hardness for Schedulability Tests for Segmented Self-Suspension

Fixed-Priority Scheduling: Suppose that the task system is scheduled by using preemptive fixed-priority scheduling. The complexity class of verifying whether the worst-case response time is no more than the relative deadline is *unknown* up to now. The evidence provided in [9] also suggests that this problem may be very difficult even for a task system with *only one self-suspending task*. The solution in [9] requires exponential time complexity for $n-1$ sporadic tasks and 1 self-suspending task. The other solutions [12][25] require pseudo-polynomial time complexity but are only sufficient schedulability tests.

The lack of something like the critical instant theorem in the ordinary sporadic task systems to reduce the search space of the worst-case behaviour has led to the complexity explosion to test exponential combinations of release patterns.

Dynamic-Priority Scheduling: The complexity class of this problem is at least co$\mathcal{NP}$-hard in the strong sense, since a special case of this problem is co$\mathcal{NP}$-complete in the strong sense [8]. It has been proved in [8] that verifying uniprocessor feasibility of sporadic tasks with constrained deadlines is strongly co$\mathcal{NP}$-complete. Therefore, when we consider constrained-deadline self-suspending task systems, the complexity class is at least co$\mathcal{NP}$-hard in the strong sense.

It is also not difficult to see that the implicit-deadline case is also at least co$\mathcal{NP}$-hard. A special case of segmented self-suspending task system is to allow a task τ_i having exactly one self-suspension interval with a *fixed* length S_i and one computation segment with WCET C_i. Therefore, the relative deadline of the computation segment of task τ_i (after it is released to be scheduled) is $D_i = T_i - S_i$. Therefore, the implicit-deadline segmented self-suspending task system is equivalent to a constrained-deadline task system, which is co$\mathcal{NP}$-complete in the strong sense. Since a special case of the problem is co$\mathcal{NP}$-complete in the strong sense, the problem is co$\mathcal{NP}$-hard in the strong sense.

6.4 Hardness for Schedulability Tests for Dynamic Self-Suspension

Fixed-Priority Scheduling: Suppose that the task system is scheduled by using preemptive fixed-priority scheduling. Similarly, with dynamic self-suspension, the complexity class of verifying whether the worst-case response time is no more than the relative deadline is *unknown* up to now. There is *no exact* schedulability analysis for this problem up to now. The solutions in [22][21][13] are only sufficient schedulability tests.

The lack of something like the critical instant theorem and the dynamics of the dynamic self-suspending behaviour have constrained the current researches to provide exact schedulability tests.

Dynamic-Priority Scheduling:

7 Rule of Thumb to Handle Self-Suspending Task Systems

8 Short Summary of the Errors and Mistakes in the State of the Art

A table to list the erratum that can be found and the reasons for the mistakes and errors.

References

1. N. Audsley, A. Burns, M. Richardson, K. Tindell, and A. Wellings. Applying new scheduling theory to static priority pre-emptive scheduling. *Software Engineering Journal*, 8(5):284–292, 1993.
2. N. C. Audsley and K. Bletsas. Fixed priority timing analysis of real-time systems with limited parallelism. In *16th Euromicro Conference on Real-Time Systems (ECRTS 2004), 30 June - 2 July 1004, Catania, Italy, Proceedings*, pages 231–238, 2004.
3. N. C. Audsley and K. Bletsas. Realistic analysis of limited parallel software / hardware implementations. In *10th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS 2004), 25-28 May 2004, Toronto, Canada*, pages 388–395, 2004.
4. K. Bletsas and N. C. Audsley. Extended analysis with reduced pessimism for systems with limited parallelism. In *11th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA 2005), 17-19 August 2005, Hong Kong, China*, pages 525–531, 2005.
5. B. Brandenburg. Improved analysis and evaluation of real-time semaphore protocols for P-FP scheduling. In *RTAS*, 2013.
6. A. Carminati, R. de Oliveira, and L. Friedrich. Exploring the design space of multiprocessor synchronization protocols for real-time systems. *Journal of Systems Architecture*, 60(3):258–270, 2014.
7. J. Chen and C. Liu. Fixed-relative-deadline scheduling of hard real-time tasks with self-suspensions. In *Proceedings of the IEEE 35th IEEE Real-Time Systems Symposium, RTSS 2014, Rome, Italy, December 2-5, 2014*, pages 149–160, 2014.
8. P. Ekberg and W. Yi. Uniprocessor feasibility of sporadic tasks with constrained deadlines is strongly conp-complete. In *27th Euromicro Conference on Real-Time Systems, ECRTS*, pages 281–286, 2015.
9. G. R. Geoffrey Nelissen, José Fonseca and V. Nelis. Timing Analysis of Fixed Priority Self-Suspending Sporadic Tasks. In *Euromicro Conference on Real-Time Systems (ECRTS)*, 2015.
10. G. Han, H. Zeng, M. Natale, X. Liu, and W. Dou. Experimental evaluation and selection of data consistency mechanisms for hard real-time applications on multicore platforms. *IEEE Transactions on Industrial Informatics*, 10(2):903–918, 2014.
11. W.-H. Huang and J.-J. Chen. Self-suspension real-time tasks under fixed-relative-deadline fixed-priority scheduling.
12. W.-H. Huang and J.-J. Chen. Schedulability and priority assignment for multi-segment self-suspending real-time tasks under fixed-priority scheduling. Technical report, Dortmund, Germany, 2015.

13. W.-H. Hung, J.-J. Chen, H. Zhou, and C. Liu. PASS: Priority assignment of real-time tasks with dynamic suspending behavior under fixed-priority scheduling. In *Design Automation Conference (DAC)*, 2015.
14. W. Kang, S. Son, J. Stankovic, and M. Amirijoo. I/O-Aware Deadline Miss Ratio Management in Real-Time Embedded Databases. In *Proc. of the 28th IEEE Real-Time Systems Symp. (RTSS)*, pages 277–287, 2007.
15. S. Kato, K. Lakshmanan, A. Kumar, M. Kelkar, Y. Ishikawa, and R. Rajkumar. RGEM: A Responsive GPGPU Execution Model for Runtime Engines. In *2011 IEEE 32nd Real-Time Systems Symposium*. Institute of Electrical & Electronics Engineers (IEEE), nov 2011.
16. H. Kim, S. Wang, and R. Rajkumar. vMPCP: a synchronization framework for multi-core virtual machines. In *RTSS*, 2014.
17. I. Kim, K. Choi, S. Park, D. Kim, and M. Hong. Real-time scheduling of tasks that contain the external blocking intervals. In *RTCSA*, pages 54–59, 1995.
18. J. Kim, B. Andersson, D. de Niz, and R. Rajkumar. Segment-fixed priority scheduling for self-suspending real-time tasks. In *Proceedings of the IEEE 34th Real-Time Systems Symposium, RTSS 2013, Vancouver, BC, Canada, December 3-6, 2013*, pages 246–257, 2013.
19. K. Lakshmanan, D. De Niz, and R. Rajkumar. Coordinated task scheduling, allocation and synchronization on multiprocessors. In *RTSS*, 2009.
20. K. Lakshmanan and R. Rajkumar. Scheduling self-suspend- ing real-time tasks with rate-monotonic priorities. In *Proceedings of the 16th IEEE Real-Time and Embedded Technology and Applications Symposium*, pages 3–12, 2010.
21. C. Liu and J. Chen. Bursty-interference analysis techniques for analyzing complex real-time task models. In *Real-Time Systems Symposium (RTSS)*, pages 173–183, 2014.
22. J. W. S. W. Liu. *Real-Time Systems*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition, 2000.
23. W. Liu, J.-J. Chen, A. Toma, T.-W. Kuo, and Q. Deng. Computation Offloading by Using Timing Unreliable Components in Real-Time Systems. In *Proceedings of the The 51st Annual Design Automation Conference on Design Automation Conference - DAC '14*. Association for Computing Machinery (ACM), 2014.
24. F. Nemati, M. Behnam, and T. Nolte. Independently-developed real-time systems on multi-cores with shared resources. In *ECRTS*, 2011.
25. J. C. Palencia and M. G. Harbour. Schedulability analysis for tasks with static and dynamic offsets. In *Proceedings of the 19th IEEE Real-Time Systems Symposium*, pages 26–37, 1998.
26. R. Rajkumar. Real-time synchronization protocols for shared memory multiprocessors. In *ICDCS*, 1990.
27. R. Rajkumar. Dealing with Suspending Periodic Tasks. Technical report, IBM T. J. Watson Research Center, 1991.
28. F. Ridouard, P. Richard, and F. Cottet. Negative Results for Scheduling Independent Hard Real-Time Tasks with Self-Suspensions. In *25th IEEE International Real-Time Systems Symposium*. Institute of Electrical & Electronics Engineers (IEEE), 2004.
29. B. Sprunt, J. P. Lehoczky, and L. Sha. Exploiting unused periodic time for aperiodic service using the extended priority exchange algorithm. In *Proceedings of the 9th IEEE Real-Time Systems Symposium (RTSS '88), December 6-8, 1988, Huntsville, Alabama, USA*, pages 251–258, 1988.

30. M. Yang, H. Lei, Y. Liao, and F. Rabee. PK-OMLP: An OMLP based k-exclusion real-time locking protocol for multi- GPU sharing under partitioned scheduling. In *DASC*, 2013.
31. M. Yang, H. Lei, Y. Liao, and F. Rabee. Improved blocking timg analysis and evaluation for the multiprocessor priority ceiling protocol. *Jounal of Computer Science and Technology*, 29(6):1003–1013, 2014.
32. M. Yang, A. Wieder, and B. Brandenburg. Global real-time semaphore protocols: A survey, unified analysis, and comparison. *to appear in 2015 RTSS*, 2015.
33. H. Zeng and M. Natale. Mechanisms for guaranteeing data consistency and flow preservation in AUTOSAR software on multi-core platforms. In *SIES*, 2011.