Task Reweighting under Global Scheduling on Multiprocessors

Aaron Block, James H. Anderson, and UmaMaheswari C. Devi
Department of Computer Science, University of North Carolina at Chapel Hill

Abstract

We consider schemes for enacting task share changes—a process called *reweighting*—on real-time multiprocessor platforms. Our particular focus is reweighting schemes that are deployed in environments in which tasks may *frequently* request *significant* share changes. Prior work has shown that fair scheduling algorithms are capable of reweighting tasks with minimal allocation error and that partitioning-based scheduling algorithms can reweight tasks with better average-case performance, but greater error. However, preemption and migration overheads can be high in fair schemes. In this paper, we consider the question of whether global scheduling techniques can improve the accuracy of reweighting relative to partitioning-based schemes and provide improved average-case performance relative to fair scheduled systems. Our conclusion is that, for soft real-time systems, global scheduling techniques provide a good mix of accuracy and average-case performance.

1 Introduction

Real-time systems that are *adaptive* in nature have received considerable recent attention [3, 9, 10, 4]. In addition, *multiprocessor* platforms are of growing importance, due to both hardware trends such as the emergence of multicore technologies and the prevalence of computationally-intensive applications for which single-processor designs are not sufficient. In prior work [3, 4], we considered the use of both fair and partitioning-based algorithms to schedule highly-adaptive workloads on (tightly-coupled) multiprocessor platforms, where the processor shares of tasks change frequently and to a significant extent. Fair scheduling techniques achieve high accuracy in enacting share changes, but do so at the expense of potentially frequent task preemptions and migrations among processors. Partitioning algorithms, in contrast, entail less overhead, but provide poorer (but sometimes acceptable) accuracy. The focus of this paper is adaptive global scheduling algorithms that avoid the high preemption and migration costs of fair scheduling techniques, yet have superior accuracy relative to partitioning-based schemes. The primary drawback of global scheduling algorithms is that, in order to fully utilize a multiprocessor system, bounded deadline misses must be acceptable. The key issue we address is whether the lower migration/preemption overheads and improved accuracy of such algorithms are sufficient to compensate for their inability to meet all deadlines.

Whisper. To motivate the need for this work, we consider two example applications under development at the University of North Carolina. The first of these is the Whisper tracking system, which performs full-body tracking in virtual environments [11]. Whisper tracks users via an array of wall- and ceiling-mounted microphones that detect white noise emitted from speakers attached to each user's hands, feet, and head. Like many tracking systems, Whisper uses *predictive techniques* to track objects. The workload on Whisper is intensive enough to necessitate a multiprocessor design. Furthermore, adaptation is required because the computational cost of making the "next" prediction in tracking an object depends on the accuracy of the previous one. Thus, the processor shares of the tasks that are deployed to implement these tracking functions will vary with time. In fact, the variance can be as much as *two orders of magnitude*. Moreover, adaptations must be enacted within *time scales as short as 10 ms.*

ASTA. The second application is the ASTA video-enhancement system [2]. ASTA is capable of improving the quality of an underexposed video feed so that objects that are indistinguishable from the background become clear and in full color. In ASTA, darker objects require more computation to correct. Thus, as dark objects move in the video, the processor shares of the tasks assigned to process different areas of the video will change. ASTA will eventually be deployed in a military-grade full-color night vision system, so tasks will need to change shares as fast as a soldier's head can turn. In the planned configuration, a 10-processor multicore platform will be used.

Dynamic sporadic tasks. In this paper, we are primarily concerned with *dynamic sporadic tasks*. Each such task T_i releases a sequence of *jobs*, T_i^1, T_i^2,... Each task is defined by the *execution cost* of each of its jobs, denoted $\mathsf{e}(T_i^j)$, and its *weight* at any time t, denoted $\mathsf{wt}(T_i, t)$, which specifies the fraction of a single processor it requires. This differs from the usual definition of a *sporadic* task, wherein per-job execution costs and weights do not change. While the terms "share," "weight," and "utilization" are often used interchangeably, we use *weight* to denote a task's desired utilization, and *share* to denote its actual guaranteed utilization. In each scheduling scheme we consider, a task's share is determined by its weight; in some of these schemes, the two are always equal, while in others, they may differ. We refer to the process of enacting task weight/share changes as *reweighting*.

Summary of results. In this paper, we consider five reweighting-capable scheduling algorithms: a previous fair algorithm developed by us called called PD^2-OF [3], which is a derivative of the PD^2 Pfair algorithm [1]; a previous partitioning-based algorithm developed by us called the *partitioned-adaptive scheduling* (PAS) algorithm [4]; the *non-preemptive-partitioned-adaptive-scheduling* (NP-PAS) algorithm, which is a non-preemptive variant of PAS; and two new algorithms proposed herein, the *changeable-earliest-deadline-first* (CNG-EDF) algorithm, which is a derivative of the well known global-earliest-deadline-first (EDF) algorithm, and the *non-preemptive-*

Scheme	Tardiness	Drift	Overload	Migrations	Preemptions
PD^2-OF	0	2	0	every quantum	every quantum
PAS	1	$\mathsf{e}_{\mathsf{max}}(T_i)$	W	weight-change events	weight-change events & job releases
NP-PAS	$\mathsf{e}(T_i^j)$ +$\mathsf{e}_{\mathsf{max}}(T_i)$+1	$\mathsf{e}_{\mathsf{max}}(T_i)$	W	weight-change events	weight-change events
CNG-EDF	$\kappa(m-1)$	$\mathsf{e}_{\mathsf{max}}(T_i)$	0	job releases	job releases
NP-CNG-EDF	$\kappa(m)$	$\mathsf{e}_{\mathsf{max}}(T_i)$	0	only in-between jobs	never

Table 1: Summary of worst-case results.

changeable-earliest-deadline-first (NP-CNG-EDF) algorithm, which is a non-preemptive variant of CNG-EDF.

Our results are summarized in Table 1, which lists the accuracy, migration cost, and preemption cost of each of the above schemes. Accuracy is assessed in terms of three quantities, "drift," "overload error," and "tardiness," which are measured in terms of the system's scheduling quantum size. *Drift* is the error, in comparison to an ideal allocation, that results due to a reweighting event [3]. (Under an ideal allocation, tasks are reweighted instantaneously, which is not possible in practice.) *Overload error*, which arises under partitioning-based schemes (see [4]), is the error that results from a scheduler's inability to allocate a task a share equal to its desired weight. *Tardiness* is the maximal amount by which any job can miss its deadline. Of these three types of error, overload error is potentially the most detrimental, since drift is a one-time error assessed per reweighting event and tardiness is bounded in the schemes we consider. Overload error, on the other hand, accumulates over time.

In Table 1, $\mathsf{e}_{\mathsf{max}}(T_i)$ denotes the *maximum execution cost* of any job of the task T_i, $\mathsf{wt}_{\mathsf{max}}(T_i)$ denotes the *maximal weight* of task T_i at any time, and W denotes the maximal weight of the $(m \cdot \lfloor 1/X \rfloor + 1)^{st}$ "heaviest" task (by maximal weight), where m is the number of processors and X is the maximal weight of the heaviest task. Furthermore,

$$\kappa(\ell) = \frac{\sum_{T_z \in \mathcal{E}_{\max}(T,\,\ell)} \mathsf{e}_{\mathsf{max}}(T_z)}{m - \sum_{T_z \in \mathcal{X}_{\max}(T,\,m-1)} \mathsf{wt}_{\mathsf{max}}(T_z)} + \mathsf{e}(T_i^j), \quad (1)$$

where $\mathcal{E}_{\max}(T, \ell)$ is the set of ℓ tasks in T with the highest *maximal* execution cost and $\mathcal{X}_{\max}(T, m-1)$ is the set of $m-1$ tasks in T with the heaviest *maximal* weight. (This bound is derived from prior work by Devi and Anderson on multiprocessor EDF scheduling [6].) Table 1 shows that algorithms that allow more frequent migrations and preemptions, like PD^2-OF, produce little drift, no overload error, and no tardiness; however, algorithms that restrict the frequency of migrations and preemptions can produce greater drift, overload error, and/or tardiness.

Contributions. Our theoretical contributions include devising CNG-EDF and NP-CNG-EDF reweighting rules, and establishing the error bounds for CNG-EDF and NP-CNG-EDF in Table 1. The question that then remains is: for the five aforementioned algorithms, how do drift, overload error, and tardiness compare to any error due to migration and preemption costs? We attempt to answer this question via extensive simulation studies of Whisper and ASTA. In these studies, real migration and preemption costs were assumed based on actual measured values. These studies confirm the expectation that, while CNG-EDF and NP-CNG-EDF provide a good compromise of accuracy and average case performance, *there exists no single "best" algorithm*: for each algorithm, application scenarios exist for which that algorithm is the best choice.

The rest of this paper is organized as follows. In Secs. 2 and 3 we discuss the CNG-EDF and NP-CNG-EDF algorithms in greater detail. Then, in Sec. 4, we establish the properties mentioned above. Our experimental evaluation is presented in Sec. 5. We conclude in Sec. 6.

2 System Model and Scheduling

In this section, we define our system model and the CNG-EDF and NP-CNG-EDF reweighting algorithms.

Sporadic task systems. We denote the ith task of a task system T as T_i (where tasks are ordered by some arbitrary method), and denote the jth job of the task T_i as T_i^j (where jobs are ordered by the sequence in which they are invoked). A *sporadic task* is defined by an *execution cost*, denoted $\mathsf{e}(T_i)$, and *weight*, denoted $\mathsf{wt}(T_i)$, which specifies the fraction of a single processor it requires. (It is customary to define a sporadic task by its execution cost and the minimum separation time between its successive jobs—we define the latter in terms of weight and execution cost below.) Fig. 1(a) depicts a one-processor system scheduled via EDF with four tasks, as defined in the figure's caption. (The other insets in the figure are considered later.) The first job of a task may be invoked or *released* at any time at or after time zero. The release time of job T_i^j is denoted $\mathsf{r}(T_i^j)$. Successive job releases of task T_i must be separated by at least $\mathsf{e}(T_i)/\mathsf{wt}(T_i)$ time. For example, in Fig. 1(a), $\mathsf{r}(T_1^1) = 0$ and $\mathsf{r}(T_1^2) = 3$. The *absolute deadline* (or just *deadline*) of job T_i^j, denoted $\mathsf{d}(T_i^j)$, equals $\mathsf{r}(T_i^j) + \mathsf{e}(T_i)/\mathsf{wt}(T_i)$. For example, in Fig. 1(a), $\mathsf{d}(T_1^1) = 3$ and $\mathsf{d}(T_1^2) = 6$. We consider a sporadic task T_i to be *active* at time t if there exists a job T_i^j (called *T_i's active job*) such that $t \in [\mathsf{r}(T_i^j), \mathsf{d}(T_i^j))$.

Dynamic sporadic task systems. A *dynamic sporadic task system* is an extension of a sporadic task system, where the weight of each task T_i is a function of time t and its execution cost can vary with each job T_i^j. We use $\mathsf{wt}(T_i, t)$ and $\mathsf{e}(T_i^j)$, respectively, to denote these two quantities. (For the remainder of the paper, whenever we refer to a "task" we are referring to a "dynamic sporadic task.") We use $\mathsf{wt}_{\mathsf{min}}(T_i)$ ($\mathsf{wt}_{\mathsf{max}}(T_i)$) to denote the *minimum* (*maximum*) *allowed weight for* T_i. As a shorthand, we use $T_i{:}[a, b]$ to denote a task T_i such that $\mathsf{wt}_{\mathsf{min}}(T_i) = a$ and $\mathsf{wt}_{\mathsf{max}}(T_i) = b$, and $T_i{:}a$ to denote $T_i{:}[a, a]$. Furthermore, we use $\mathsf{e}_{\mathsf{max}}(T_i)$ to denote the maximal execution cost of any job of T_i. Fig. 1(b) gives an example.

For dynamic sporadic tasks, the *absolute deadline* of a job T_i^j equals $\mathsf{r}(T_i^j) + \mathsf{e}(T_i^j)/\mathsf{wt}(T_i, \mathsf{r}(T_i^j))$. In the absence of reweighting, consecutive job releases ($\mathsf{r}(T_i^j)$ and $\mathsf{r}(T_i^{j+1})$) of a task T_i must be separated by at least $\mathsf{e}(T_i^j)/\mathsf{wt}(T_i, \mathsf{r}(T_i^j))$. For example,

in Fig. 1(b), $\mathsf{r}(T_1^2) - \mathsf{r}(T_1^1) = 2/(1/3) = 6$, $\mathsf{r}(T_1^3) - \mathsf{r}(T_1^2) = 2/(1/2) = 4$, and $\mathsf{d}(T_1^3) = 10 + 1/(1/2) = 12$.

A task T_i *changes weight* or *reweights* at time t if $\mathsf{wt}(T_i, t-\epsilon) \neq \mathsf{wt}(T_i, t)$ where $\epsilon \to 0^+$. If a task T_i changes weight at a time t_c between the release and the deadline of some job T_i^j, then the following three actions *may* occur:

- The execution cost of T_i^j *may* be reduced to the amount of time for which T_i^j has executed prior to t_c.
- $\mathsf{r}(T_i^{j+1})$ *may* be less than $\mathsf{r}(T_i^j) + \mathsf{e}(T_i^j)/\mathsf{wt}(T_i, \mathsf{r}(T_i^j))$.
- If T_i^{j+1} is released before $\mathsf{r}(T_i^j) + \mathsf{e}(T_i^j)/\mathsf{wt}(T_i, \mathsf{r}(T_i^j))$, then since $\mathsf{d}(T_i^j) = \mathsf{r}(T_i^j) + \mathsf{e}(T_i^j)/\mathsf{wt}(T_i, \mathsf{r}(T_i^j))$, jobs T_i^j and T_i^{j+1} will "overlap." (In the variant of the sporadic model defined earlier, every job's deadline is at or before its successors's release.) Hence, we say that a job T_i^j is *active* at time t iff $t \in [\mathsf{r}(T_i^j), \mathsf{min}(\mathsf{r}(T_i^{j+1}), \mathsf{d}(T_i^j)))$.

The reweighting rules we present in Sec. 3 state under what conditions the above actions occur and by how much before $\mathsf{r}(T_i^j) + \mathsf{e}(T_i^j)/\mathsf{wt}(T_i, \mathsf{r}(T_i^j))$ the job T_i^{j+1} can be released. Since a reweighting event may cause a job's execution cost to decrease, we introduce the notion of a job T_i^j's *actual execution cost*, denoted $\mathsf{ae}(T_i^j)$, which represents the total amount of execution time that T_i^j will receive.

When a task reweights, there can be a difference between when it "initiates" the change and when the change is "enacted." The time at which the change is *initiated* is a user-defined time; the time at which the change is *enacted* is dictated by a set of conditions discussed shortly. We use the *scheduling weight of a task T_i at time t*, denoted $\mathsf{swt}(T_i, t)$, to represent the "last enacted weight of T_i". Formally, $\mathsf{swt}(T_i, t)$ equals $\mathsf{wt}(T_i, u)$, where u is the last time at or before t that a weight change was enacted for T_i. It is important to note that, *henceforth, we compute task deadlines and releases using scheduling weights.*

Scheduling. Under both CNG-EDF and NP-CNG-EDF, "ready" jobs are prioritized by deadline, with earlier deadlines having higher priority. ("Ready" will be formally defined shortly.) Deadline ties are resolved arbitrarily, but consistently. Under CNG-EDF, an arriving job with higher priority preempts the executing job with the lowest priority if no processor is available. The preempted job may later resume execution on a different processor. Under NP-CNG-EDF, the arriving job waits until some job completes execution and a processor becomes available. Thus, under NP-CNG-EDF, once scheduled, a job is guaranteed execution until completion without interruption. Fig. 1(b) depicts a CNG-EDF schedule of the task system T described above, and Fig. 1(c) depicts a NP-CNG-EDF schedule of the same system.

For an arbitrary scheduling algorithm $\mathcal{A}$ and an arbitrary task system T, we let $\mathcal{S}$ denote an m-processor schedule $\mathcal{A}$ of T, and let $\mathsf{A}(\mathcal{S}, T_i^j, t_1, t_2)$ denote the total time allocated to T_i^j in $\mathcal{S}$ in $[t_1, t_2)$. Similarly, we use $\mathsf{A}(\mathcal{S}, T_i, t_1, t_2)$ and $\mathsf{A}(\mathcal{S}, T, t_1, t_2)$, respectively, to denote the total time allocated to all jobs of T_i in $\mathcal{S}$ and all tasks of T in $\mathcal{S}$, over the interval $[t_1, t_2)$. We say that the value of $\mathsf{A}(\mathcal{S}, T_i^j, 0, t)$ is the amount that T_i^j has *executed by* t. For example in Fig. 1(b), $\mathsf{A}(\mathcal{S}, T_1^1, 0, 6) = 2$, $\mathsf{A}(\mathcal{S}, T_1^1, 0, 12) = 2$, and $\mathsf{A}(\mathcal{S}, T_1^1, 3, 12) = 0$.

Definition 1 (Halted). As discussed later, if a reweighting event in schedule $\mathcal{S}$ occurs at time t, then it is possible that some job T_i^j is *halted* at t. In this case, $\mathsf{ae}(T_i^j)$ is set to $\mathsf{A}(\mathcal{S}, T_i^j, 0, t)$.

Definition 2 (Completed). If $\mathcal{S}$ is an m-processor CNG-EDF or NP-CNG-EDF schedule of the task system T, then a job $T_i^j \in T$ is said to have *completed by time t in $\mathcal{S}$* iff T_i^j has executed for $\mathsf{e}(T_i^j)$ by t in $\mathcal{S}$, or T_i^j has halted by time t. A task T_i is said to have *completed at time t in $\mathcal{S}$* if at time t every job of T_i that has been released by t has completed. A task T_i is said to have *entirely completed by time t* iff *all* jobs of T_i in T have completed. For example, in Fig. 1(b), T_1^1 has completed by time 3, T_1 is complete (but *not* entirely complete) at time 3 but not complete at time 6, and T_4 is entirely complete at time 4.

Definition 3 (Pending and Ready). For an arbitrary scheduling algorithm $\mathcal{A}$, if $\mathcal{S}$ is an m-processor schedule of the task system T under $\mathcal{A}$, then a job T_i^j is said to be *pending at time t in $\mathcal{S}$* if $\mathsf{r}(T_i^j) \leq t$ and T_i^j is not complete by t in $\mathcal{S}$. For example, in Fig. 1(a), the job T_2^1 is pending over the range $[0, 9)$. Note that a job can be pending, but not active, if it misses its deadline. A pending job T_i^j is said to be *ready at time t in $\mathcal{S}$* if all prior jobs of task T_i have completed by t. For example, in Fig. 1(a), the job T_2^1 is ready over the range $[0, 9)$. A job T_i^j can be pending but not ready if T_i^{j-1} has not completed by $\mathsf{r}(T_i^j)$.

3 Task Reweighting

We now introduce two new reweighting rules that are CNG-EDF extensions of the PD^2-OF reweighting rules presented by us previously [3]. As mentioned before, these rules work by modifying future job release times and deadlines. At the end of this section, we discuss how to adjust these rules for NP-CNG-EDF.

For simplicity, we assume that the actual execution cost for any job is equal to its specified execution cost, *unless* a task reweights while a job is active. Then *and only then* can the actual execution cost of a job be less than its execution cost. (This assumption can be removed at the expense of more complicated notation.) In this scenario, the actual execution cost of the job is determined by the rules we present shortly.

Let T be a task system in which some task T_i initiates a weight change from weight w to weight v at time t_c. Let $\mathcal{S}$ be the m-processor CNG-EDF schedule of T. Let T_i^j be the active job of T_i at t_c. If $\mathsf{e}(T_i^j) - \mathsf{A}(\mathcal{S}, T_i^j, 0, t) > 0$, then let $\mathsf{rem}(T_i^j, t_c) = \mathsf{e}(T_i^j) - \mathsf{A}(\mathcal{S}, T_i^j, 0, t)$; otherwise, $\mathsf{rem}(T_i^j, t_c) = \mathsf{e}(T_i^{j+1})$. Note that $\mathsf{rem}(T_i^j, t_c)$ denotes the actual remaining computation in T_i's current job or the size of T_i's next job if the current job has completed. The *deviance of job T_i^j of task T_i at time t* is defined as $\mathsf{dev}(T_i^j, t) = \int_{\mathsf{r}(T_i^j)}^{t} \mathsf{swt}(T_i, u)du - \mathsf{A}(\mathcal{S}, T_i^j, 0, t)$. The choice of which rule to apply depends on whether deviance is positive or negative. If positive, then we say that T_i is *positive-changeable at time t_c from weight w to v*; otherwise T_i is *negative-changeable at time t_c from weight w to v*. Because T_i initiates its weight change at t_c, $\mathsf{wt}(T_i, t_c) = v$ holds; however, T_i's scheduling weight does not change until the weight change has been *enacted*, as specified in the rules below. Note that if t_c occurs between the initiation and enaction of a previous reweighting event of T_i,

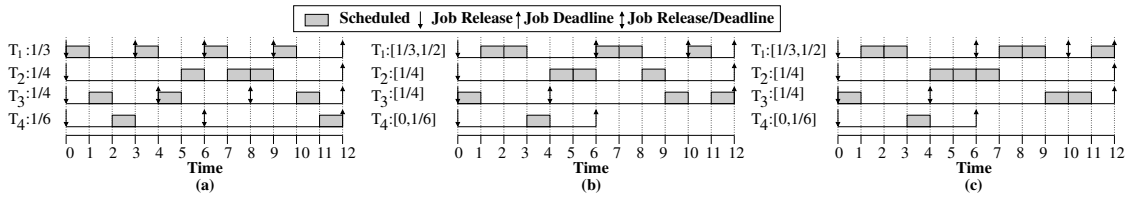

Figure 1: A one-processor (sporadic or dynamic sporadic) system T with four tasks. Inset **(a)** depicts an EDF schedule T where the tasks are defined as follows: T_1 with weight 1/3 and $\mathsf{e}(T_1) = 1$, T_2 with weight 1/4 and $\mathsf{e}(T_2) = 3$, T_3 with weight 1/4 and $\mathsf{e}(T_3) = 1$, and T_4 with weight 1/6 and $\mathsf{e}(T_4) = 1$. In insets **(b)** and **(c)**, the tasks are defined as follows: T_1 has an initial weight of 1/3 and increases to 1/2 at time 6, $\mathsf{e}(T_1^1) = \mathsf{e}(T_1^2) = 2$, and $\mathsf{e}(T_1^3) = 1$; T_2 has a constant weight of 1/4 and $\mathsf{e}(T_2^1) = 3$; T_3 has a constant weight of 1/4, $\mathsf{e}(T_3^1) = 1$, and $\mathsf{e}(T_3^2) = 2$; and T_4 has an initial weight of 1/6 and decrease to 0 at time 6 (*i.e.*, T_4 "leaves" the system at 6) and $\mathsf{e}(T_4^1) = 1$. Inset (b) depicts a CNG-EDF schedule of T. Inset (c) depicts a NP-CNG-EDF schedule of T. All ties are broken in favor of the task with the lower index.

then the previous event is skipped, *i.e.*, treated as if it had not occurred. As discussed later, any "error" associated with skipping a reweighting event like this is accounted for when determining drift.

Rule P: If T_i is positive-changeable at time t_c from weight w to v, then one of two actions is taken: **(i)** if $\mathsf{d}(T_i^j) > \mathsf{rem}(T_i, t_c)/v$, then T_i^j is halted, its weight change is enacted, and a new job of size $\mathsf{rem}(T_i, t_c)$ is issued for it with a release time of t_c; **(ii)** otherwise, its weight change is enacted at time $\mathsf{d}(T_i^j)$, *i.e.*, the scheduling weight does not change until the end of the current job.

Rule N: If T_i is negative-changeable at time t_c from weight w to v, then one of two actions is taken: **(i)** if $v > w$, then T_i^j is halted, its weight change is enacted, and a new job of size $\mathsf{rem}(T_i, t_c)$ is issued for it with a release time equal to the time t at which $\mathsf{dev}(T_i^j, t) = 0$ holds; **(ii)** otherwise, the weight change is enacted at time $\mathsf{d}(T_i^j)$.

Intuitively, Rule P changes a task's weight by halting its current job and issuing a new job of size $\mathsf{rem}(T_i, t_c)$ with the new weight if doing so would improve its deadline. A (one-processor) example of a positive-changeable task is given in Fig. 2(a). (We discuss the terms drift, IDEAL allocations, and SW allocations in Sec. 4.) The depicted example consists of a task system T with four tasks as defined in the figure's caption. Note that, since T_2, T_3, and T_4 have the same deadline, we have arbitrarily chosen T_4 to have the lowest priority. In inset (a), T_4 is positive-changeable since at time 2 it has not yet been scheduled. Note that halting T_4's current job and issuing a new job of size one improves T_4's scheduling priority, *i.e.*, $\mathsf{d}(T_4^1) = 6 > \frac{7}{2} = \mathsf{d}(T_4^2)$. Notice that the second job of T_4 is issued 6/4 quanta after time 2. This spacing is in keeping with a new job of weight 4/6 issued at time 2.

Rule N changes the weight of a task by one of two approaches: **(i)** if a task *increases* its weight, then Rule N adjusts the release time of its next job so that it is commensurate with the new weight; **(ii)** if a task *decreases* its weight, then Rule N waits until the end of the current job and then issues the next job with a deadline that is commensurate with the new weight. A (one-processor) example of a negative-changeable task that increases its weight is given in Fig. 2(b). The depicted example consists of the same tasks as in (a), except that we have chosen T_4 to have the highest priority. Notice that the second job of T_4 is issued at time 3, which is the time such that $\mathsf{dev}(T_4, 3) = \int_0^3 \mathsf{swt}(T_i, u)du - \mathsf{A}(\mathcal{S}, T_4, 0, 3) = 1 - 1 = 0$. Recall that the deadline (release time) of the ith ($(i+1)^{th}$) job of a task T_j is given by $\mathsf{r}(T_i^j) + \mathsf{e}(T_i^j)/(\mathsf{swt}(T_i, \mathsf{r}(T_i^j)))$. Hence, if a task T_i of weight v were to issue a job of size $y = \mathsf{A}(\mathcal{S}, T_i^j, 0, t_c) - \mathsf{dev}(T_i^j, t)$ at time t_c, then the release time of its next job would be $t_c + y/v$. A (one-processor) example of a negative-changeable task that decreases its weight is given in Fig. 2(c). The depicted example consists of the same four tasks except that T_4 has an initial weight of 4/6 and decreases its weight at time 1, and T_1 joins the system as soon as T_4's weight change is enacted.

Since these rules change the ordering of a task in the priority queues that determine scheduling, the time complexity for reweighting one task is $O(logN)$, where N is the number of tasks in the system.

Modifications for NP-CNG-EDF. In order to adapt the rules P and N to work for NP-CNG-EDF, the only modification we need to make is when these rules are initiated. If a task reweights *before or after* the active job has been scheduled, then the rules P and N are initiated as before. (Note that if the active job *has not* been scheduled, then its deviance is positive, and if the active job *has* been scheduled, its deviance is negative.) However, if a task changes its weight while the active job T_i^j is executing, then the initiation of the weight change is delayed *until* T_i^j *has completed* or T_i^j *is no longer active*, whichever is first. Note that when a task T_i changes its weight from u to v at time t_c in NP-CNG-EDF, then $\mathsf{wt}(T_i, t_c) = v$ holds, regardless of whether the initiation of rule P or N must be delayed.

4 Tardiness and Drift Bounds

In this section, we formally present and prove tardiness and drift bounds for the CNG-EDF algorithm. Because any set of reweighting rules will cause the "actual" schedule to deviate from the "ideal" schedule, the tardiness bounds reflect CNG-EDF's accuracy at *scheduling* the job-set created by CNG-EDF. The drift bounds, on the other-hand, reflect CNG-EDF's accuracy at creating a job-set that mimics the "ideal" task system, where weight changes can always be initiated and enacted instantaneously. To this end, we introduce two new theoretical scheduling algorithms: the *scheduling-weight processor-sharing* (SW) scheduling algorithm and the *ideal processor-sharing* (IDEAL) scheduling algorithm. Both algorithms have the ability to preempt and swap tasks at arbitrarily small intervals. However, SW allocates each task a share equal to its *scheduling weight*; moreover, SW *will not allocate* capacity to a task if its active job has received an allocation equal to its actual execution cost. IDEAL, on the other hand, al-

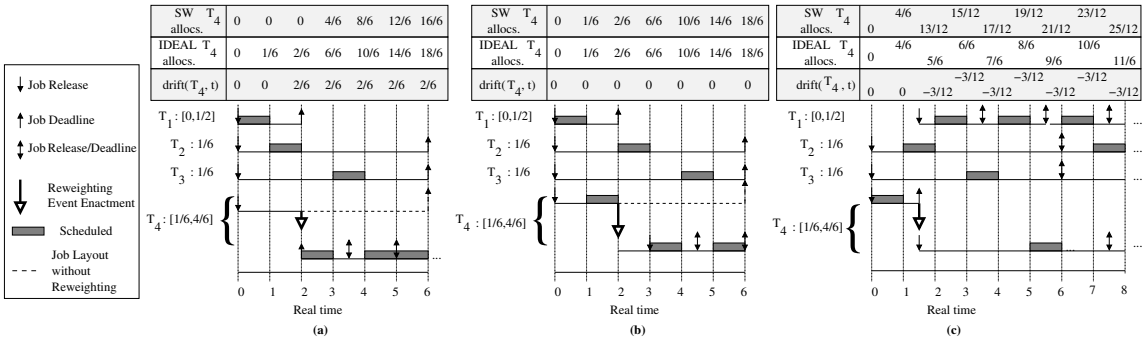

Figure 2: A one-processor system consisting of four tasks, T_1:$[0, 1/2]$, T_2:$1/6$, T_3:$1/6$, and T_4:$[1/6, 4/6]$, where the execution cost of every job is one. The dotted lines represent the interval up to T_4's next deadline, which due to reweighting has been changed (as indicated by the solid arrow). The drift, allocations in IDEAL, and allocations in SW for T_4 are labeled as a function of time across the top. **(a)** The CNG-EDF schedule for the scenario where T_1 is in the system initially and leaves at time 2, T_4 has an initial weight of 1/6 that increases to 4/6 at time 2, and T_4 has the lowest scheduling priority. Since T_4 is not scheduled by time 2, it has positive deviance and changes its weight via Rule P, causing T_4^1 to be halted, and T_4^2 to be releaed at 2 deadline of 9/2, and its drift to become 2/6. **(b)** The same scenario as in (a) except that T_4 has higher priority than both T_2 and T_3. Since T_4 has been scheduled by time 2, it has negative deviance and changes its weight via Rule N, causing its next job to have a release time of 3 while maintaining a drift of zero. **(c)** T_1 joins the system at time 6/4 and T_4 has an initial weight of 4/6 that decreases to 1/6 at time 1. Since T_4 has negative deviance at time 1, it is changed via Rule N, causing T_4's next job to have a deadline of 15/2 and T_4 to have a drift of $-3/12$.

locates each task a share equal to its *weight* at each instant; and, unlike SW, IDEAL *will not stop allocating* capacity to a task *unless* that task has received an allocation equal to the total execution cost of all of its jobs. (For simplicity, we have assumed that every job in T is released as early as possible. This assumption can be removed at the cost of more complex notation. If we did not make this assumption, then the allocation function for IDEAL would equal zero between active jobs.) We provide below a more in-depth explanation of these two algorithms.

4.1 Tardiness and Lag

We begin by defining the SW scheduling algorithm.

The SW scheduling algorithm. In order to establish tardiness bounds for CNG-EDF, we compare allocations produced by CNG-EDF to those produced by SW. Under SW, at each instant t, each non-complete job of each task T_i is allocated a fraction of a processor equal to $\mathsf{swt}(T_i, t)$. Furthermore, we consider SW to be "clairvoyant" in the sense that SW can use the value of $\mathsf{ae}(T_i^j)$ to determine if T_i^j has completed before it has halted. More specifically, for any schedule $\mathcal{S}W$ under SW of any task system T, we say that T_i^j has *completed by time t in $\mathcal{S}W$* iff T_i^j has executed for $\mathsf{ae}(T_i^j)$ by t.

For example, consider the one-processor task system T depicted in Fig. 3. Inset (a) depicts a CNG-EDF schedule and insets (b) depicts T's SW schedule. Notice that in the SW schedule T_1 does not receive any allocations over the interval $[3, 6)$. This is because at time 3 the total allocation to T_1^1 in the SW schedule equals $\mathsf{ae}(T_1^1) = 1$, hence, T_1^1 is complete at time 3. However, at time 6, T_1^2 is released, and therefore T_1 has an incomplete job with a scheduling weight of 1/2. Hence, T_1 begins to receive allocations equal to its scheduling weight, which is now 1/2. Note that we assume that every job release, deadline, execution cost, and actual execution cost for a SW schedule to be the same as that in CNG-EDF.

Lag. If $\mathcal{S}$ is an m-processor schedule under CNG-EDF of the task system T and $\mathcal{S}W$ is an m-processor schedule under SW of the same task system T, then the *lags at time t of a job T_i^j, task T_i, and task system T*, respectively, are defined by Eqns. (2)–(4).

$$\mathsf{lag}(T_i^j, t) = \mathsf{A}(\mathcal{S}W, T_i^j, 0, t) - \mathsf{A}(\mathcal{S}, T_i^j, 0, t) \quad (2)$$

$$\mathsf{lag}(T_i, t) = \mathsf{A}(\mathcal{S}W, T_i, 0, t) - \mathsf{A}(\mathcal{S}, T_i, 0, t) \quad (3)$$

$$\mathsf{LAG}(T, t) = \mathsf{A}(\mathcal{S}W, T, 0, t) - \mathsf{A}(\mathcal{S}, T, 0, t) \quad (4)$$

Note that $\mathsf{LAG}(T, t) = \sum_{T_i \in T} \mathsf{lag}(T_i, t)$. The lag of a job (or task or system) represents by how much a job (or task or system) is under/over-allocated compared to the SW schedule at time t. For example, in Fig. 3, $\mathsf{lag}(T_3^1, 1) = 1/4 - 0 = 1/4$, $\mathsf{lag}(T_3^1, 2) = 2/4 - 1 = -1/2$, $\mathsf{lag}(T_1^1, 2) = 2/3 - 0 = 2/3$, $\mathsf{lag}(T_1^1, 3) = 3/3 - 0 = 1$, and $\mathsf{lag}(T_1^1, 6) = 3/3 - 1 = 0$.

4.2 Tardiness Proof

In prior work, Devi and Anderson [6] proved that in any m-processor EDF schedule of a *sporadic* task system T (where the total weight of all tasks is at most m) the tardiness of each job of any task T_i is at most $\kappa(m-1)$ where, $\kappa(m-1)$ is as defined in (1). Their proof consists primarily of three lemmas/theorems: **(i)** if the LAG of T is bounded in the m-processor EDF schedule $\mathcal{S}$ of T, then tardiness is bounded; **(ii)** the LAG of T in $\mathcal{S}$ is bounded; **(iii)** by (i) and (ii), the tardiness of each job of any task T_i in T is at most $\kappa(m-1)$.

Since Devi and Anderson were proving tardiness bounds for a sporadic task system, they were able to utilize the fact that a job T_i^j and its successor T_i^{j+1} do not "overlap," *i.e.*, $\mathsf{d}(T_i^j) \leq \mathsf{r}(T_i^{j+1})$ holds for any sporadic task T_i. However, this property can be weakened without affecting their proof (barring some minor notational changes), so that their proof can be adapted to prove tardiness bounds for a *dynamic* sporadic task system. Specifically, the Devi and Anderson proof can be used to show that the tardiness of CNG-EDF is bounded by $\kappa(m-1)$. If the following properties hold.

(W) $\sum_{T_i \in T} \mathsf{wt}(T_i, t) \leq m$ for all t.

(V) For any job T_i^j and its successor T_i^{j+1}, if $\mathsf{d}(T_i^j) > \mathsf{r}(T_i^{j+1})$,

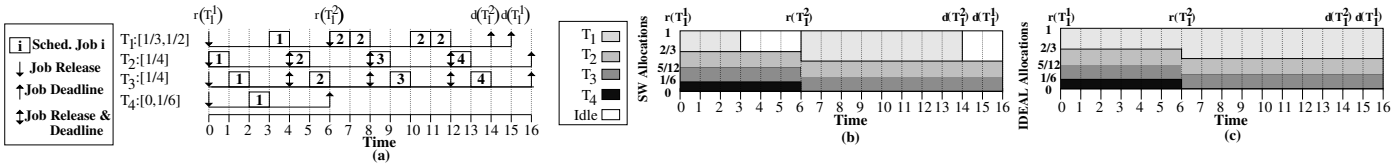

Figure 3: A one-processor task system T with four tasks: $\mathsf{e}(T_1^1) = 5$ and T_1 has an initial weight of 1/3 that increases to 1/2 via case (i) of rule P at time 6 and as a result, $\mathsf{ae}(T_1^1) = 1$, $\mathsf{e}(T_1^2) = 4$, $\mathsf{r}(T_1^2) = 6$ and $\mathsf{d}(T_1^2) = 14$; T_2 has a constant weight of 1/4 and a constant execution cost of 1; T_3 has a constant weight of 1/4 and a constant execution cost of 1; and T_4 has an initial weight of 1/6 that decreases to 0 at time 6 and $\mathsf{e}(T_4^1) = 1$. Inset **(a)** depicts the CNG-EDF schedule of T. The number in each box denotes which job is scheduled, *e.g.*, over the range $[0, 1)$, T_2^1 is executing, and over the range $[4, 5)$, T_2^2 is executing. Inset **(b)** depicts the SW schedule of T. Note that since $\mathsf{ae}(T_1^1) = 1$ at time 3, T_1^1 is complete in SW, *i.e.*, T_1 receives no allocations under SW over the range $[3, 6)$. Inset **(c)** depicts the IDEAL schedule of T. Note that the IDEAL allocation to any task T_i equals $\int_{t_1}^{t_2} \mathsf{wt}(T_i, u)du$ if T_i is active over the range $[t_1, t_2)$. For insets (b) and (c), the releases and deadlines of the jobs T_1^1 and T_1^2 are depicted.

then T_i^j, must have completed before $\mathsf{r}(T_i^{j+1})$ in both the CNG-EDF and SW schedules of T.

Since (W) can be easily satisfied, for the remainder of this subsection, we show that CNG-EDF satisfies property (V). (Unfortunately, due to space constraints, we are not able to present the Devi and Anderson proof with the necessary (minor) adjustments in the body of this paper. Therefore, we have placed this proof in an appendix to this paper, which can be found on the author's web-page at `http://www.cs.unc.edu/~anderson/papers.html`.) In order to show that property (V) holds, we show that for any job T_i^j in an arbitrary dynamic sporadic task system T, if $\mathsf{d}(T_i^j) > \mathsf{r}(T_i^{j+1})$, then T_i^j must have completed before $\mathsf{r}(T_i^{j+1})$ in both the CNG-EDF and SW schedules of T. To this end, let $\mathcal{S}$ be the m-processor CNG-EDF schedule of some dynamic task system T, where $\sum_{T_i \in T} \mathsf{wt}(T_i, t) \leq m$ for all t, and let $\mathcal{SW}$ be the m-processor SW schedule of the same task system.

Lemma 1. *For a task T_i, if $\mathsf{r}(T_i^{j+k}) < \mathsf{d}(T_i^j)$, where $j, k \geq 1$, then T_i^j will have completed by $\mathsf{r}(T_i^{j+k})$ in $\mathcal{S}$ and $\mathcal{SW}$.*

Proof. Suppose that $\mathsf{r}(T_i^{j+k}) < \mathsf{d}(T_i^j)$ holds. By the definition of $\mathsf{d}(T_i^j)$, the minimum separation between job releases, and rules P and N, $\mathsf{r}(T_i^{j+k}) < \mathsf{d}(T_i^j)$ holds only if T_i reweighted and halted while T_i^j was active. Without loss of generality, let t_c be the earliest such time. Then, by the rules P and N, $\mathsf{r}(T_i^{j+k}) \geq t_c$. Hence, T_i^j will have halted and thus completed by $\mathsf{r}(T_i^{j+k})$ in $\mathcal{S}$.

It remains to be shown that T_i^j will have completed by $\mathsf{r}(T_i^{j+k})$ in $\mathcal{SW}$. Since T_i^j is halted at t_c, it must be the case that T_i changed its weight via case (i) of rule P or N at t_c. However, both cases follow easily by the clairvoyant nature of $\mathcal{SW}$. □

Modifications for NP-CNG-EDF. In NP-CNG-EDF, if a job is released and is ready at time t, and the newly-released job has a deadline that is earlier than some other job executing at t, the newly released job cannot preempt the lower-priority job. If no processor is available at t, then this will lead to a *priority inversion*. In such a scenario, the waiting ready, higher-priority job is referred to as a *blocked job*, and the executing lower-priority job is referred to as a *blocking job*. A task T_i is said to be *blocked* at t if T_i is not executing at t and the earliest pending job (*i.e.*, the ready job) of T_i has a higher priority than at least one job executing at t. For example, in Fig. 1(c), T_2 is the blocking job over the interval $[6, 7)$, and T_1 is the blocked job over the same interval.

The major difference between Devi and Anderson's tardiness-bound proof for EDF and NP-EDF for sporadic tasks is that in their NP-EDF proof they calculate the upper bound on the length of time for which a task can be blocked. Because of the blocking factor, the bound they construct for NP-EDF is $\kappa(m)$. As before, Devi and Anderson in their proof for NP-EDF rely on the property of sporadic tasks that consecutive jobs do not "overlap." And, as before, this requirement can be weakened without affecting their proof, so that their proof holds for a *dynamic* sporadic task set, so long as conditions (V) and (W) hold. Since the reweighting rules for NP-CNG-EDF are essentially the same as the reweighting rules for CNG-EDF, by Lem. 1, conditions (V) and (W) hold for any m-processor NP-CNG-EDF schedule, of any task set T so long as $m \leq \sum_{T_i \in T} \mathsf{wt}(T_i, t)$ holds for all t. (As before, due to space constraints we are forced to present the Devi and Anderson proof, with its modification, in its entirety in an appendix found on the author's web page.) Hence, NP-CNG-EDF's tardiness bound is $\kappa(m)$.

4.3 Drift

We now turn our attention to the issue of measuring "drift" under CNG-EDF. In order to measure the "drift" of a task system T, we compare the SW schedule of T to that of an "ideal" reweighting scheme that enacts reweighting changes instantaneously. Under the *ideal processor sharing* (IDEAL) scheduling algorithm, at each instant t, each task T_i in T is allocated a share equal to its weight $\mathsf{wt}(T_i, t)$. Hence, if $\mathcal{I}$ is the IDEAL schedule of T, then over the interval $[t_1, t_2)$, the task T_i is allocated $\mathsf{A}(\mathcal{I}, T_i^j, t_1, t_2) = \int_{t_1}^{t_2} \mathsf{wt}(T_i, u)du$ time. As we mentioned earlier, IDEAL is similar to SW, with two major exceptions: **(i)** under IDEAL, each task receives an allocation equal to its *weight*, whereas under SW, each task receives an allocation equal to its *scheduling weight*; and **(ii)** under IDEAL, a task does not stop receiving allocations unless its total allocation equals the total execution cost of all of its jobs, whereas under SW, a task will stop receiving allocations if its active job has received an allocation equal to its actual execution cost. For example, consider the IDEAL schedule of the task system T depicted in Fig. 3(c). Notice that, over the range $[3, 6)$, the task T_1 receives allocations equal to its weight at every instant. Compare this to the SW schedule (inset (b)), in which T_1 receives *no* allocations over the range $[3, 6)$.

For most real-time scheduling algorithms, the difference between the ideal and actual allocations a task receives lies within

some bounded range centered at zero. For example, under a *uniprocessor* EDF (*i.e.*, CNG-EDF without weight changes) schedule, the difference between the ideal and actual allocations for a task lies within $(-\mathsf{e}_{\mathsf{max}}(T_i), \mathsf{e}_{\mathsf{max}}(T_i))$. When a weight change occurs, the same bounds are maintained except that they may be centered at a different value. For example, in Fig. 2(a), the range is originally $(-1, 1)$, but after the reweighting event, it is $(-4/6, 8/6)$. This lost allocation is called *drift*. Given this loss (barring further reweighting events) T_i's drift will not change. In general, a task's drift per reweighting event will be non-negative (non-positive) if it increases (decreases) its weight. Under CNG-EDF, the drift of a task T_i at time t is defined as

$$\mathsf{drift}(T_i, t) = \mathsf{A}(\mathcal{I}, T_i^j, 0, u) - \mathsf{A}(\mathcal{S}W, T_i^j, 0, u), \qquad (5)$$

where $\mathcal{S}W$ is the schedule of T under SW, $\mathcal{I}$ is the schedule of T under IDEAL, and u is the last time a reweighting event of T_i was enacted before t.

Theorem 1. *The absolute value of the per-event drift under* CNG-EDF *for each task T_i is less than* $\mathsf{e}_{\mathsf{max}}(T_i)$.

Proof Sketch. If a task T_i changes its weight at time t_c via rule P, then when this weight change is enacted at time t_e (*i.e.*, at t_c under case (i) or at $\mathsf{d}(T_i^j)$ under case (ii)), then it is as though allocation equal to $\mathsf{A}(\mathcal{I}, T_i^j, \mathsf{r}(T_i^j), t_e) - \mathsf{A}(\mathcal{S}W, T_i^j, \mathsf{r}(T_i^j), t_e)$ is "lost." For example in Fig. 2(a), the task T_4 "loses" an allocation of 2/6. Since this value (per reweighting event) is always less than $\mathsf{e}_{\mathsf{max}}(T_i)$, the absolute value of drift is less than $\mathsf{e}_{\mathsf{max}}(T_i)$.

If a task T_i changes its weight at time t_c via rule N, and T_i decreases its weight (case (ii)), then the weight change will be enacted at $\mathsf{d}(T_i^j)$. Since the maximum allocation T_i can receive in $\mathcal{S}W$ during T_i^j is $\mathsf{e}_{\mathsf{max}}(T_i)$, $\mathsf{A}(\mathcal{S}W, T_i^j, t_c, \mathsf{d}(T_i^j)) - \mathsf{A}(\mathcal{I}, T_i^j, t_c, \mathsf{d}(T_i^j)) \leq \mathsf{e}_{\mathsf{max}}(T_i)$. Thus, the absolute value of the drift incurred is at most $\mathsf{e}_{\mathsf{max}}(T_i)$. For example, in Fig. 2(c), the drift incurred by T_4 is $-3/12$, *i.e.*, $\mathsf{drift}(T_4, t) = -3/12$, where $t \geq 3/2$. If T_i increases its weight (case (i)), then it incurs zero drift, since it *immediately* enacts the weight change (*i.e.*, the scheduling weight changes immediately). Hence, the absolute value of the drift incurred by this reweighting event is less than $\mathsf{e}_{\mathsf{max}}(T_i)$. For example, in Fig. 2(b), the drift incurred by T_4 is 0, *i.e.*, $\mathsf{drift}(T_4, t) = 0$, where $t \geq 2$. □

Modifications for NP-CNG-EDF. Note that delaying the initiation of a reweighting event does not substantially increase the drift incurred per reweighting event, since the longest a reweighting event can be delayed is the execution cost of the active job. If T_i^j is the active job of T_i at t_c, and if T_i's reweighting event is delayed until some time t, then at t either **(i)** T_i^j has a non-positive deviance (*i.e.*, T_i^j completes before its deadline), or **(ii)** T_i^j is not active at t (*i.e.*, T_i^j does not complete before its deadline, and thus is not active at t). In either case, the active job (if it exists) is negative-changeable. Hence, if the task increases its weight, then the only drift the task will incur for this reweighting event results from delaying the initiation of its reweighting event, *i.e.*, at most $\mathsf{e}_{\mathsf{max}}(T_i)$. If T_i decreases its weight, then delaying the reweighting event will not affect drift, since the enactment of the reweighting event would occur at $\mathsf{d}(T_i^j)$ regardless of whether the initiation of the reweighting event was delayed or not.

5 Experimental Results

The results of this paper are part of a longer-term project on adaptive real-time allocation in which both Whisper and ASTA described earlier, will be used as test applications. In this section, we provide extensive simulations of Whisper and ASTA as scheduled by PD^2-OF, PAS, NP-PAS, CNG-EDF, and NP-CNG-EDF.

Whisper. As noted earlier, Whisper tracks users via speakers that emit white noise attached to each user's hands, feet, and head. Microphones located on the wall or ceiling receive these signals and a tracking computer calculates each speaker's position by measuring signal delays. Whisper is able to compute the time-shift between the transmitted and received versions of the sound by performing a *correlation* calculation on the most recent set of samples. By varying the number of samples, Whisper can trade measurement accuracy for computation—with more samples, the more accurate and more computationally intensive the calculation. As a signal becomes weaker, the number of samples is increased to maintain the same level of accuracy. As the distance between a speaker and microphone increases, the signal strength decreases. This behavior (along with the use of predictive techniques mentioned in the introduction) can cause task-share changes of up to two orders of magnitude every 10ms. Since Whisper continuously performs calculations on incoming data, at any point in time, it does not have a significant amount of "useful" data stored in cache. As a result, migration/preemption costs in Whisper are fairly small (at least, on a tightly-coupled system, as assumed here, where the main cost of a migration is a loss of cache affinity). In addition, fairness and real-time guarantees are important due to the inherent "tight coupling" among tasks required to accurately perform triangulation calculations.

ASTA system. Before describing ASTA in detail, we review some basics of videography. All video is a collection of still images called *frames*. Associated with each frame is an *exposure time*, which denotes the amount of time the camera's shutter was open while taking that frame. Frames with faster exposure times capture moving objects with more detail, while frames with slower exposure times are brighter. If a frame is *underexposed* (*i.e.*, the exposure time is too fast), then the image can be too dark to discern any object. The ASTA system can correct underexposed video while maintaining the detail captured by faster exposure times by combining the information of multiple frames. To intuitively understand how ASTA achieves this behavior, consider the following example. If a camera, **A**, has an exposure time of $1/30^{th}$ of a second, and a second camera, **B**, has an exposure time of $1/15^{th}$ of a second, then for every two frames shot by camera **A** the shutter is open for the same time as one frame shot by **B**. ASTA is capable of exploiting this observation in order to allow camera **A** to shoot frames with the detail of $1/30^{th}$ of a second exposure time but the brightness of $1/15^{th}$ of a second exposure time. As noted earlier, darker objects require more computation than lighter objects to correct. Thus, as dark objects move in the video, the processor shares of tasks assigned to process different areas of the video will change. As a result, tasks will need to adjust their weights as quickly as an object can move

across the screen. Since ASTA continuously performs calculations based on previous frames, it performs best when a substantial amount of "useful" data is stored in the cache. As a result, migration/preemption costs in ASTA are fairly high. In addition, while strong real-time and fairness guarantees would be desirable in ASTA, they are not as important here as in Whisper, because tasks can function more independently in ASTA.

Experimental system set up. Unfortunately, at this point in time, it is not feasible to produce experiments involving a real implementation of either Whisper or ASTA, for several reasons. First, both the existing Whisper and ASTA systems are single-threaded (and non-adaptive) and consist of several thousands of lines of code. All of this code has to be re-implemented as a multi-threaded system, which is a nontrivial task. Indeed, because of this, it is *essential* that we first understand the scheduling and resource-allocation trade-offs involved. The development of PD^2-OF, PAS, NP-PAS, CNG-EDF, and NP-CNG-EDF can be seen as an attempt to articulate these tradeoffs. Additionally, the focus of this paper is on scheduling methods that facilitate adaptation—we have *not* addressed the issue of devising mechanisms for determining *how* and *when* the system should adapt. Such mechanisms will be based on issues involving virtual-reality and multimedia systems that are well beyond the scope of this paper. For these reasons, we have chosen to evaluate the schemes discussed in this paper via simulations of Whisper and ASTA. While just simulations, most of the parameters used here were obtained by implementing and timing the scheduling algorithms discussed in this paper and some of the signal-processing and video-enhancement code in Whisper and ASTA, respectively, on a real multiprocessor testbed. Thus, the behaviors in these simulations should fairly accurately reflect what one would see in a real Whisper or ASTA implementation.

For both Whisper and ASTA, the simulated platform was assumed to be a shared-memory multiprocessor, with four 2.7-GHz processors and a 1-ms quantum. All simulations were run 61 times. Both systems were simulated for 10 secs. (Note that longer simulations return similar results.) We implemented and timed each scheduling scheme considered in our simulations on an actual testbed that is the same as that assumed in our simulations, and found that all scheduling and reweighting computations could be completed within 5μs. We considered this value to be negligible in comparison to a 1-ms quantum and thus did not consider scheduling overheads in our simulations. For both Whisper and ASTA, we conducted two types of experiments: **(i)** all preemption and migration costs were the same and corresponded to a loss of cache affinity; and **(ii)** the preemption cost was set to some value and the migration cost was varied. If a task was preempted and then migrated, we assumed that it incured the maximum of the two costs. We ignored the issue of bus contention, since in prior work, Holman and Anderson have shown that bus contention can be virtually eliminated in Pfair-scheduled systems by *staggering* quantum allocations on different processors [7]. Staggering would be trivial to apply in PAS and NP-PAS as well, since in PAS, processors run nearly independently of each other. Furthermore, since CNG-EDF and NP-CNG-EDF are event-based rather than quantum-based, jobs are unlikely to begin executing simultaneously. Based on measurements taken on our testbed system, we estimated Whisper's migration cost as 2μs–10μs, and ASTA's as 50μs–60μs. While we believe that these costs may be typical for a wide range of systems, in our experiments we varied the preemption/migration cost over a slightly larger range. For all experiments, the maximum execution cost of PAS and NP-PAS was 7ms and 5ms for CNG-EDF and NP-CNG-EDF. These values were determined by profiling each system beforehand to determine the "best" compromise of accuracy and performance.

While the ultimate metric for determining the efficacy of both systems would be user perception, this metric is not currently available, for reasons discussed earlier. Therefore, we compared each of the tested schemes by comparing against allocations in the IDEAL algorithm. In particular, we measured both the "average under-allocation" and "fairness factor" for each task set at the end of each simulation (*i.e.*, 10 secs.). The *average under-allocation* (UA) is the average amount each task is behind its IDEAL allocation (this value is defined to be nonnegative, *i.e.*, for a task that is not behind its IDEAL, this value is zero). The *fairness factor* (FF) of a task set is the largest deviance from the allocations in IDEAL between any two tasks (*e.g.*, if a system has three tasks, one that deviates from its IDEAL allocation by -10, another by 20, and the third by 50, then the FF is $50 - (-10) = 60$). The FF is a good indication of how fairly a scheme allocates processing capacity. A lower FF means the system is more fair. For applications like Whisper, where the output generated by multiple tasks is periodically combined, a low FF is important, since if any one task is "behind," then performance of the entire system is impacted; however, for applications like ASTA, where tasks are more independent, a high FF does not effect the system performance nearly as much. These metrics should provide us with a reasonable impression of how well the tested schemes will perform when Whisper and ASTA are fully re-implemented.

Whisper experiments. In our Whisper experiments, we simulated three speakers (one per object) revolving around pole in a 1m $\times$ 1m room with a microphone in each corner, as shown in Fig. 4. The pole creates potential occlusions. One task is required for each speaker-microphone pair, for a total of 12 tasks. In each simulation, the speakers were evenly distributed around the pole at an equal distance from the pole, and rotated around the pole at the same speed. The starting position for each speaker was set randomly. As mentioned above, as the distance between a speaker and microphone changes, so does the amount of computation necessary to correctly track the speaker. This distance is (obviously) impacted by a speaker's movement, but is also lengthened when an occlusion is caused by the pole. The range of weights of each task was determined (as a function of a tracked object's position) by implementing and timing the basic computation of the correlation algorithm (an

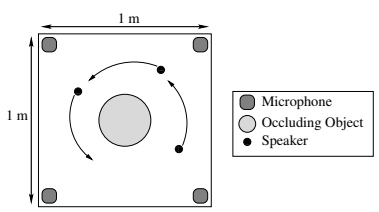

Figure 4: The simulated Whisper system.

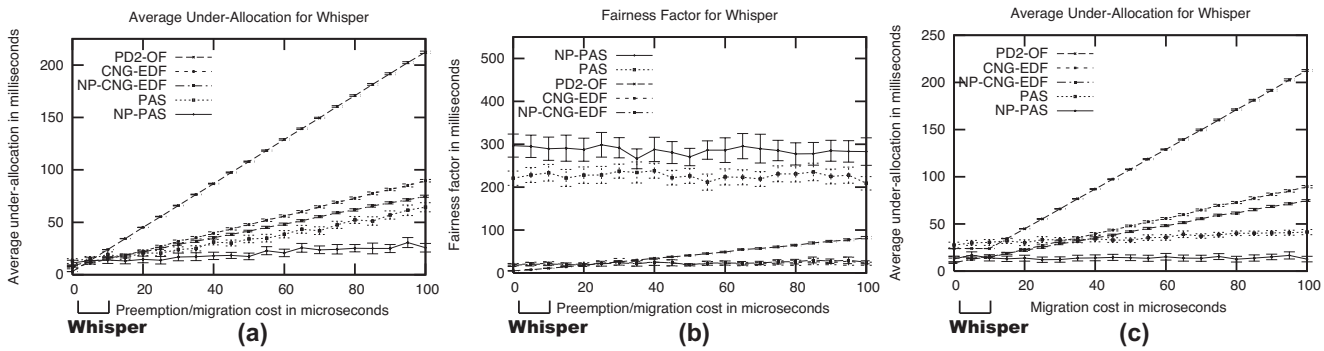

Figure 5: **(a)** The average under-allocation (UA) and **(b)** the fairness factor (FF) for Whisper as a function of preemption/migration cost, and **(c)** the average UA for Whisper as a function of migration cost (preemption cost is fixed at 10μs), as scheduled by each tested algorithm. The key in each graph is in the order that the schemes appear in that graph at 100μs. 98% confidence intervals are shown. Note that in (b), CNG-EDF and NP-CNG-EDF are indistiguible from each-other.

accumulate-and-multiply operation) on our testbed system.

In the Whisper simulations, we made several simplifying assumptions. First, all objects are moving in only two dimensions. Second, there is no ambient noise in the room. Third, no speaker can interfere with any other speaker. Fourth, all objects move at a constant rate. Fifth, the weight of each task changes only once for every 5cm of distance between its associated speaker and microphone. Sixth, all speakers and microphones are omnidirectional. Finally, all tasks have a minimum weight based on measurements from our testbed system and a maximum weight of 1.0. A task's current weight at any time lies between these two extremes and depends on the corresponding speaker's current position. Even with theses assumptions, frequent share adaptations are required.

We conducted Whisper experiments in which the tracked objects were sampled at a rate of 1,000 Hz, the distance of each object from the room's center was set at 50cm, the speed of each object was set at 5 m/sec. (this is within the speed of human motion), and the maximum execution cost, migration, and preemption cost were varied. However, due to page limitations, the graphs below are a representative sampling our collected data.

The first set of graphs in Fig. 5 show the result of the Whisper simulations conducted to compare PD^2-OF, PAS, NP-PAS, CNG-EDF, and NP-CNG-EDF. Insets (a) and (b) depict the average UA and FF, respectively, for each scheme, where the preemption cost is varied from 0 to 100μs and the migration cost equals the preemption cost. Inset (c) depicts the average UA for each scheme, where the preemption cost is set at 10μs (the maximum expected preemption cost for Whisper) and the migration cost is varied from 0 to 100μs. There are five things worth noting here. First, when the preemption/migration cost is varied over the range 2 to 10μs, the UA is about the same for all schemes (inset (a)); however, PD^2-OF has the best FF (inset (b)). Second, while CNG-EDF and NP-CNG-EDF do not have the best UA for the expected preemption/migration costs for Whisper, for higher preemption/migration costs, *i.e.*, preemption/migration costs larger than 10μs, CNG-EDF and NP-CNG-EDF both have a substantially better UA than PD^2-OF and better FF than either PAS or NP-PAS. Third, as the migration cost (but not preemption cost) of a task increases, the UA of PAS and NP-PAS increases slowly (inset (c)). However the performance of the other three schemes decays quickly. Fourth, the confidence intervals for the FF for CNG-EDF, NP-CNG-EDF, and PD^2-OF are smaller than for PAS and NP-PAS, since CNG-EDF, NP-CNG-EDF, and PD^2-OF have better accuracy. Fifth, in inset (c), PD^2-OF and CNG-EDF's UA do not appreciably increase until the migration cost exceeds 10μs, this is because until the migration cost is 10μs PD^2-OF and CNG-EDF incur the maximum of the migration or preemption cost, which is 10μs.

ASTA experiments. In our ASTA experiments, we simulated a 640×640-pixel video feed where a grey square that is 160×160 pixels moves around in a circle with a radius of 160 pixels on a white background. This is illustrated in Fig. 6. The grey square makes one complete rotation every ten seconds. The position of the grey square on the circle is random. Each frame is divided into sixteen 160×160-pixel regions; each of these regions is corrected by a different task. A task's weight is determined by whether the grey square covers its region. By analyzing ASTA's code, we determined that the grey square takes three times more processing time to correct than the white background. Hence, if the grey square completely covers a task's region, then its weight is three times larger than that of a task with an all-white region. The video is shot at a rate of 25 frames per second, and as a result, each frame has an exposure time of 40ms.

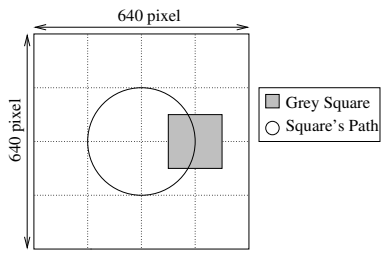

Figure 6: The simulated ASTA system.

The second set of graphs, in Fig. 7, show the result of the ASTA simulations conducted to compare the five scheduling algorithms. Insets (a) and (b) depict the average UA and FF, for each scheme, where the preemption cost is varied from 0 to 100μs and the migration cost equals the preemption cost. Inset (c) depicts the average UA for each scheme, where the preemption cost is set at 60μs (the maximum expected preemption cost for ASTA) and the migration cost is varied from 0 to 100μs. There are two things worth noting here. First, when the preemption/migration cost is varied over the range 50 to 60μs, NP-PAS and PAS have the smallest UA (inset (a)); however, CNG-EDF and NP-CNG-EDF both

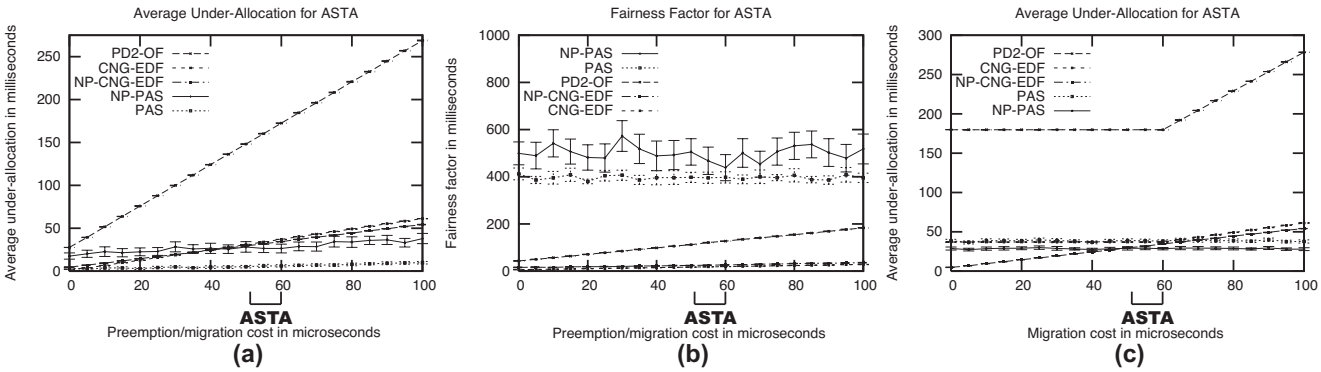

Figure 7: **(a)** The average under-allocation (UA) and **(b)** the fairness factor (FF) for ASTA as a function of preemption/migration cost, and **(c)** the average UA for ASTA as a function of migration cost (preemption cost is fixed at 60μs), as scheduled by each tested algorithm. The key in each graph is in the order that the schemes appear in that graph at 100μs. 98% confidence intervals are shown. Note that in (b), CNG-EDF and NP-CNG-EDF are indistiguible from each-other.

Scheme	Provides Hard Real-Time Guarantees	Has Low Migration/ Preemption Costs	Provides Strong Fairness Guarantees
PD^2-OF	✓		✓
(NP-)PAS		✓	
(NP-)CNG-EDF		✓	✓

Table 2: Summary of algorithm performance.

have an UA that is competitive with both PAS and NP-PAS (inset (a)) and have a *substantially* smaller FF (inset (b)). Second, in inset (c) PD^2-OF and CNG-EDF's UA do not appreciably increase until the migration cost equals 60μs. This occurs for the same reason that PD^2-OF and CNG-EDF did not noticeably increase, until 10μs in Fig. 5(c).

6 Concluding Remarks

We have presented a two new multiprocessor reweighting schemes, CNG-EDF and NP-CNG-EDF, which reduce migration costs and preemptions at the expense of allowing deadline misses. We have also presented both analytical and experimental comparisons of these schemes with a more accurate but more migration-prone scheme, PD^2-OF, and two less accurate partitioning schemes that have lower tardiness, PAS and NP-PAS. These results suggest that when it is critical that every task make its deadline and migration/preemption costs are low (*i.e.*, systems like Whisper), then PD^2-OF is the best choice; when preemption/migration costs are high (*i.e.*, either Whisper or ASTA as implemented on a system where the processors are not as tightly integrated), average case performance is of the utmost importance, and fairness and timeless are less important, then either PAS or NP-PAS may be the best choice; and when preemption/migration costs are high and and a good mix of average-case performance and fairness factor is beneficial (*i.e.*, systems like ASTA), then either CNG-EDF or NP-CNG-EDF may be the best choice. Thus, *each algorithm is of value* and will be the best choice in certain application scenarios, as is summarized in Table 2.

While our focus in this paper has been on scheduling techniques that *facilitate* fine-grained adaptations, techniques for determining *how* and *when* to adapt are equally important. Such techniques can either be application-specific (*e.g.*, adaptation policies unique to a tracking system like Whisper) or more generic (*e.g.*, feedback-control mechanisms incorporated within scheduling algorithms [9]). Both kinds of techniques warrant further study, especially in the domain of multiprocessor platforms.

References

[1] J. Anderson and A. Srinivasan. Mixed Pfair/ERfair scheduling of asynchronous periodic tasks. *Journal of Computer and System Sciences*, 68(1):157–204, 2004.

[2] E. Bennett and L. McMillan. Video enhancement using per-pixel virtual exposures. *ACM Trans. on Graphics*, 24(3):845–852, 2005.

[3] A. Block, J. Anderson, and G. Bishop. Fine-grained task reweighting on multiprocessors. In *Proc. of the 11th IEEE Int'l Conf. on Embedded and Real-Time Comp. Sys. and Apps.*, pages 429–35, 2005.

[4] A. Block and J. Anderson. Accuracy verus migration overheads in multiprocessor reweighting. Algorithms. In submission.

[5] J. Carpenter, S. Funk, P. Holman, A. Srinivasan, J. Anderson, and S. Baruah. A categorization of real-time multiprocessor scheduling problems and algorithms. In Joseph Y. Leung, editor, *Handbook on Scheduling Algorithms, Methods, and Models*, pages 30.1–30.19. Chapman Hall/CRC, Boca Raton, Florida, 2004.

[6] U. Devi and J. Anderson. Tardiness bounds under global EDF scheduling on a multiprocessor. In *Proc. of the 26th IEEE Real-time Sys. Symp.*, pages 330–41, 2005.

[7] P. Holman and J. Anderson. Implementing Pfairness on a symmetric multiprocessor. In *Proc. of the 10th IEEE Real-time and Embedded Technology and App. Symp.*, pages 544–553, 2004.

[8] J. Lopez, J. Diaz, and D. Garcia. Utilization bounds for EDF scheduling on real-time multiprocessor systems. *Real-Time Sys.*, 28(1):39–68, 2004.

[9] C. Lu, J. Stankovic, G. Tao, and S. Son. Design and evaluation of a feedback control EDF scheduling algorithm. In *Proc. of the 20th IEEE Real-time Sys. Symp.*, pages 44–53, 1999.

[10] I. Stoica, H. Abdel-Wahab, K. Jeffay, S. Baruah, J. Gehrke, and C.G. Plaxton. A proportional share resource allocation algorithm for real-time, time-shared systems. In *Proc. of the 17th IEEE Real-time Sys. Symp.*, pages 288–299, 1996.

[11] N. Vallidis. *WHISPER: A Spread Spectrum Approach to Occlusion in Acoustic Tracking*. PhD thesis, University of North Carolina, Chapel Hill, North Carolina, 2002.