## به نام پروردگار یگانه

### پروژهی سیستمهای بیدرنگ

Real-Time Systems

دانشكدهي مهندسي كامييوتر

مدرس: دکتر صفری



### گزارش پروژه ۲ فاز اول

ميترا قلي پور ۴۰۱۱۰۶۳۶۳

مليكا عليزاده ٢٠١١٠۶٢٥٥

## موضوع: تحلیل تأثیر تجاوز از زمان اجرای اسمی (NET) بر زمان پاسخ در سامانههای بی درنگ

#### مقدمه:

با توجه به این موضوع که در سامانههای بیدرنگ رعایت صحت زمانی چالشی اساسی است و به دلیل پیچیدگی سختافزارهای مدرن، تعیین دقیق زمان اجرای بدترین حالت WCET دشوار بوده و به جای آن، زمان اجرای اسمی NET که بر اساس اندازه گیریهای تجربی تعیین می شود، مورد استفاده قرار می گیرد. از آنجایی که تجاوز از NET می تواند منجر به افزایش غیرخطی زمان پاسخ و نقض مهلتهای زمانی شود در این پروژه سعی داریم به تحلیل اثرات تجاوز از NET در سامانههای بی درنگ چندهسته ای ناهمگن و ارائه روشی سیستماتیک برای شناسایی و بررسی این اثررات بپردازیم.

در این پروژه روش پیشنهادی را با سیاستهای مختلف زمانبندی مثل RM, EDF, FIFO و در مدلهای -RM و در مدلهای -Preemptive, Non و در این پروژه روش پیشنهادی را با سیاستهای مختلف زمانبندی مثل Partitioned Scheduling, Global Scheduling و تحت شرایط Preemptive

همچنين تجاوز از NET مدلسازي شده و اثرات آن تا زمان از بين رفتن پيامدهاي آن بررسي ميشود. براي هر مقدار تجاوز e كران بالايي زمان پاسخ (Ri(e به عنوان يک رويداد تجاوز محاسبه شده و مقادير بحراني e با استفاده از الگوريتم هاي جستجوي نمايي و دودويي شناسايي ميشوند.

### پیادهسازی

در این بخش ابتدا کتابخانه های لازم را اضافه میکنیم تا ابزارهای مورد نیاز برای تحلیل سیستم های بی درنگ فراهم شود.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from dataclasses import dataclass
from typing import List, Tuple, Dict, Optional
from enum import Enum
import random
import math
from scipy.optimize import minimize_scalar
import warnings
```

از Enum برای مشخص کردن نوع سیاست زمانبندی و مدلهای زمانبندی و مدل زمانبندی استفاده میکنیم.

```
class SchedulingPolicy(Enum):
    RM = "Rate Monotonic"
    EDF = "Earliest Deadline First"
    FIFO = "First In First Out"

class SchedulingType(Enum):
    PREEMPTIVE = "Preemptive"
    NON_PREEMPTIVE = "Non-Preemptive"

class SchedulingModel(Enum):
    PARTITIONED = "Partitioned"
    GLOBAL = "Global"
```

ساختار کلاس task، یک وظیفهی بی درنگ را با تمام پارامترهای لازم برای تحلیل نمایش می دهد. این پارامترها شامل id برای هر تسک، دورهی زمانی تکرار وظیفه، زمان اسمی اجرا، مهلت نسبی، اولویت وظیفه، زمان رسیدن و شناسهی هستهای که وظیفه به آن اختصاص داده شده است، می شود.

```
@dataclass
class Task:
    id: int
    period: float
    net: float
    deadline: float
    priority: int
    arrival_time: float = 0.0
    core_id: int = 0
```

کلاس RealTimeSystem سیستم را با پارامترهای قابل پیکربندی مقداردهی اولیه میکند و از سیستمهای چندهستهای ناهمگن پشتیبانی مینماید. بهصورت پیشفرض، سیستم شامل ۴ هسته با زمانبندی RM و به صورت preemptive و partitioned است. این کلاس عملکردهای اصلی زیر را پیادهسازی میکند:

- تولید و مدیریت وظایف
- تحلیل زمان پاسخ با در نظر گرفتن NET
  - شناسایی نقاط بحرانی
  - تحلیل زمانبندیپذیری
- پشتیانی از Partitioned Scheduling, Global Scheduling

در تابع generate\_task\_set مجموعه وظایف برای تحلیل تولید می شود. از الگوریتم UUniFast استفاده می شود تا توزیع یکنواخت بهرهوری تضمین شود، که برای مدل سازی واقع گرایانهی سیستم های بی درنگ اهمیت دارد. این متد از تعداد وظایف متغیر (بین ۱۰ تا ۵۰۰) پشتیبانی میکند. هر وظیفه دارای موارد زیر خواهد بود:

- دورهی زمانی تصادفی در بازهی مشخص شده
- NET محاسبه شده برای رسیدن به بهرهوری هدف
  - ضرب العجل برابر با دوره
- اولویت اختصاص یافته بر اساس سیاست زمانبندی

پارامترها: تعداد وظایف برای تولید، بهرهوری هدف سیستم، بازهی دورههای وظایف خروجی: لیستی از وظایف تولیدشده

```
def generate_task_set(self, num_tasks: int, target_utilization: float,
                     period_range: Tuple[int, int] = (10, 100)) -> List[Task]:
    utilizations = self._uunifast(num_tasks, target_utilization)
    for i in range(num_tasks):
       period = random.randint(period_range[0], period_range[1])
        net = utilizations[i] * period
       deadline = period
        if self.scheduling_policy == SchedulingPolicy.RM:
        elif self.scheduling_policy == SchedulingPolicy.EDF:
           priority = deadline
            period=period,
            deadline=deadline,
            priority=priority,
        tasks.append(task)
    if self.scheduling_model == SchedulingModel.PARTITIONED:
        self._assign_tasks_to_cores_partitioned(tasks)
```

الگوریتم UUniFast بهرهوری وظایف را بهصورت یکنواخت تولید میکند، بهگونهای که مجموع آنها برابر با بهرهوری هدف باشد. این موضوع برای ایجاد مجموعه وظایف واقعگرایانه با نرخهای بهرهوری مختلف است. پارامترها: تعداد مقادیر بهرهوری برای تولید و مجموع هدف بهرهوریها خروجی: لیستی از مقادیر بهرهوری

```
def _uunifast(self, n: int, target_u: float) -> List[float]:
    if n == 1:
        return [target_u]

    utilizations = []
    sum_u = target_u

    for i in range(n - 1):
        next_sum_u = sum_u * (random.random() ** (1.0 / (n - i)))
         utilizations.append(sum_u - next_sum_u)
        sum_u = next_sum_u

    utilizations.append(sum_u)
    return utilizations
```

در تابع assign\_tasks\_to\_cores\_partitioned\_ اختصاص وظایف به هسته ها برای زمانبندی پارتیشن شده با استفاده از Decreasing، Worst-Fit کی استفاده از استفاده از الگوریتم های تخصیص وظیفه انجام می شود. همچنین استراتژی های مختلف تخصیص را پیاده سازی می کند: Round-Robin Decreasing، First-Fit Decreasing، Best-Fit پارامتر: لیستی از وظایف برای تخصیص به هسته ها

```
def _assign_tasks_to_cores_partitioned(self, tasks: List[Task]) -> None:
   self.core_assignments = {i: [] for i in range(self.num_cores)}
   core_utilizations = [0.0] * self.num_cores
    tasks_by_util = sorted(tasks, key=lambda t: t.net/t.period, reverse=True)
    for task in tasks_by_util:
       task_utilization = task.net / task.period
       best core = 0
       max_remaining_capacity = 1.0 - core_utilizations[0]
        for core_id in range(1, self.num_cores):
            remaining_capacity = 1.0 - core_utilizations[core_id]
            if remaining_capacity > max_remaining_capacity:
                max_remaining_capacity = remaining_capacity
        if core_utilizations[best_core] + task_utilization <= 1.0:</pre>
            task.core_id = best_core
            self.core_assignments[best_core].append(task.id)
            core_utilizations[best_core] += task_utilization
            min_util_core = core_utilizations.index(min(core_utilizations))
            task.core_id = min_util_core
            self.core_assignments[min_util_core].append(task.id)
            core_utilizations[min_util_core] += task_utilization
```

در تابع calculate\_response\_time\_with\_overrun کران بالای زمان پاسخ در صورت تجاوز از NET محاسبه می شود. و آنالیز زمان پاسخ را برای هر دو حالت Partitioned Scheduling, Global Scheduling انجام می دهد. این پیاده سازی شامل زمان اجرای خود وظیفه، تداخل ناشی از وظایف با اولویت بالاتر، الگوهای تداخل متفاوت برای زمان بندی و محاسبه تا رسیدن به همگرایی می شود.

رین کر و مقدار e باید تحلیل شود و مقدار e خروجی: کران بالای زمان پاسخ

```
def calculate_response_time_with_overrun(self, task_index: int, overrun_amount: float) -> float:
    if task_index >= len(self.tasks):
        return float('inf')

    task = self.tasks[task_index]

    if self.scheduling_model == SchedulingModel.PARTITIONED:
        return self._calculate_response_time_partitioned(task_index, overrun_amount)
    else:
        return self._calculate_response_time_global(task_index, overrun_amount)
```

در تابع calculate\_response\_time\_partitioned\_ زمان پاسخ برای Partitioned محاسبه می شود. به این شکل که هر وظیفه به یک هسته خاص اختصاص داده می شود و تنها با وظایف دیگر روی همان هسته رقابت می کند.

```
def _calculate_response_time_partitioned(self, task_index: int, overrun_amount: float) -> float:
    task = self.tasks[task_index]
    same_core_tasks.sort(key=lambda t: t.priority)
    task_pos = next(i for i, t in enumerate(same_core_tasks) if t.id == task.id)
    R = task.net + overrun_amount
    R_prev = 0
    max_iterations = 100
    iteration = 0
    while abs(R - R_prev) > 0.001 and iteration < max_iterations:
       R_{prev} = R
        interference = 0
        for j in range(task_pos):
            higher_priority_task = same_core_tasks[j]
            if self.scheduling_policy == SchedulingPolicy.RM:
                num_preemptions = math.ceil(R / higher_priority_task.period)
                task_execution = higher_priority_task.net + (overrun_amount * 0.1)
                interference += num_preemptions * task_execution
            elif self.scheduling_policy == SchedulingPolicy.EDF:
                if higher_priority_task.deadline <= task.deadline:</pre>
                    num_preemptions = math.ceil(R / higher_priority_task.period)
                    interference += num_preemptions * higher_priority_task.net
            elif self.scheduling_policy == SchedulingPolicy.FIF0:
                if higher_priority_task.arrival_time <= task.arrival_time:
                    interference += higher_priority_task.net
        R = task.net + overrun_amount + interference
        if R > task.deadline:
```

در تابع calculate\_response\_time\_global رمان پاسخ برای calculate\_response\_time\_global محاسبه می شود. به این شکل که وظایف می توانند روی هر هسته ای اجرا شوند که منجر به تحلیل تداخل پیچیده تری می شود با استفاده از m پردازنده که m نشان دهنده تعداد هسته است.

```
def _calculate_response_time_global(self, task_index: int, overrun_amount: float) -> float:
    task = self.tasks[task_index]
    R = task.net + overrun_amount
    R_prev = 0
   max_iterations = 100
    iteration = 0
    while abs(R - R_prev) > 0.001 and iteration < max_iterations:</pre>
        R_{prev} = R
        total_interference = 0
        higher_priority_workload = 0
        for j in range(task_index):
            higher_priority_task = self.tasks[j]
            if self.scheduling_policy == SchedulingPolicy.RM:
                if higher_priority_task.period < task.period:</pre>
                    workload = math.ceil(R / higher_priority_task.period) * higher_priority_task.n
                    higher_priority_workload += workload
            elif self.scheduling_policy == SchedulingPolicy.EDF:
                if higher_priority_task.deadline <= task.deadline:</pre>
                    workload = math.ceil(R / higher_priority_task.period) * higher_priority_task.n
                    higher_priority_workload += workload
            elif self.scheduling_policy == SchedulingPolicy.FIFO:
                if higher_priority_task.arrival_time <= task.arrival_time:</pre>
                    higher_priority_workload += higher_priority_task.net
        max_parallel_interference = max(0, higher_priority_workload - (self.num_cores - 1) * R)
        total_interference = min(higher_priority_workload, max_parallel_interference + (self.num_c
        overrun_interference = overrun_amount * 0.2 * (task_index + 1) / len(self.tasks)
        total_interference += overrun_interference
        R = task.net + overrun_amount + total_interference / self.num_cores
```

در تابع find\_critical\_overrun\_values مقادیر بحرانی که باعث افزایش غیرخطی زمان پاسخ میشوند شناسایی شده و برای این کار از دو روش استفاده میشود: جستجوی نمایی برای شناسایی اولیه بهصورت تقریبی و جستجوی دودویی برای یافتن دقمة، نقاط بحرانی

یات اندیس وظیفهای که باید تحلیل شود و بیشترین مقدار e مورد بررسی و اندازه گام برای تحلیل خروجی: لیستی از زوجهای (مقدار بحرانی، زمان پاسخ) برای نقاط بحرانی

در تابع زیر سربار مهاجرت وظایف در زمانبندی global محاسبه میشود. وظایف میتوانند بین هسته ها جابه جا شوند که این موضوع باعث ایجاد سربارهایی مانند از دسترفتن کش، همگامسازی حافظه و موارد مشابه می شود. پارامترها: تعداد وظایف موجود در سیستم خروجی: برآوردی از سربار مهاجرت به صورت درصدی از زمان اجرا

```
def _estimate_migration_overhead(self, num_tasks: int) -> float:
   base_overhead = 0.05
   complexity_factor = min(num_tasks / 100.0, 1.0)
   return base_overhead * (1 + complexity_factor)
```

در تابع analyze\_system\_performance عملکرد سیستم تحت شرایط مختلف بررسی می شود. این متد نیازمندی های اصلی تحلیل را پیاده سازی می کند که شامل تغییر تعداد وظایف، نرخهای مختلف بهرهوری، محاسبه زمانهای پاسخ و شناسایی نقاط بحرانی، تولید شاخصهای عملکرد و نمودارهای بصری پارامترها: لیستی از تعداد وظایف برای تحلیل، لیستی از نرخهای بهرهوری برای آزمون خروجی: دیکشنری شامل نتایج تحلیل

در تابع زیر مقایسه عملکرد بین مدلهایPartitioned Scheduling, Global Scheduling انجام می شود. این متد مجموعه وظایف یکسانی را تحت هر دو مدل زمان بندی تحلیل می کند تا مزایا و معایب هر مدل را بررسی کرده و مشخص کند که در چه شرایطی هرکدام عملکرد بهتری دارند.

پارامترها: لیستی از تعداد وظایف برای تحلیل، لیستی از نرخهای بهرهوری برای آزمون خروجی: دیکشنری شامل نتایج مقایسه

### نتایح مقایسه در ماتریس مقایسه نمایش داده می شود.

در دو تابع print\_detailed\_results, generate\_performance\_graphs نمودارها و خروجی دقیق برای تحلیل نتایج ایجاد می شود. نمودارها شامل زمان پاسخ در برابر تعداد وظایف، زمان پاسخ در برابر بهرهوری، تحلیل زمانبندی پذیری، تحلیل تأثیر تجاوز NET می شود. خروجی شامل خلاصه ای از پیکربندی ها، کران های زمان پاسخ، نقاط بحرانی و تحلیل زمانبندی پذیری می شود.

# برای بررسی باید ویژگیهای سامانه بیدرنگ را در تابع main مشخص نماییم.

```
Real-Time System NET Overrun Analysis
Scheduling Model: Partitioned
  - Schedulable: True
- Avg Response Time: 1.71
  - Max Response Time: 3.89
  - Schedulable: True
- Avg Response Time: 8.12
  - Avg Response Time: 1.63
- Max Response Time: 4.53
  - Avg Response Time: 4.72
  - Max Response Time: 11.77
   - Avg Response Time: 1.45
    Avg Response Time: 2.48
  - Schedulable: True
   - Avg Response Time: 3.98
  - Schedulable: True
- Avg Response Time: 1.20
  - Schedulable: True
```

```
- Avg Response Time: 2.50
  - Avg Response Time: 3.61
  - Schedulable: True
  - Avg Response Time: 1.26
  - Schedulable: True
  - Avg Response Time: 2.39
  - Max Response Time: 7.24
  - Schedulable: True
- Avg Response Time: 3.98
DETAILED REAL-TIME SYSTEM ANALYSIS RESULTS
  - Scheduling Policy: Rate Monotonic
- Scheduling Type: Preemptive
    Scheduling Model: Partitioned
DETAILED RESULTS BY CONFIGURATION:
  Maximum Response Time: 3.895
Configuration: 10 tasks, utilization = 0.50 (Partitioned Scheduling)
Schedulable: True
  Average Response Time: 5.371
  Tasks with Critical Overrun Points: 0
```

```
Configuration: 50 tasks, utilization = 0.25 (Partitioned Scheduling)
  Average Response Time: 1.632
 Maximum Response Time: 4.526
  Tasks with Critical Overrun Points: 12
Configuration: 50 tasks, utilization = 0.50 (Partitioned Scheduling)
  Average Response Time: 3.191
 Maximum Response Time: 8.983
  Average Response Time: 1.448
  Core Load Distribution: [23, 25, 26, 26]
Configuration: 100 tasks, utilization = 0.75 (Partitioned Scheduling)
  Maximum Response Time: 10.519
  Core Load Distribution: [24, 26, 25, 25]
  Schedulable: True
  Average Response Time: 1.198
  Maximum Response Time: 4.469
  Core Load Distribution: [50, 49, 51, 50]
```

