Conclusion and Advanced Topics

Prof. Dr. Jian-Jia Chen

LS 12, TU Dortmund

14, July, 2015

Things You May Already Forget

- Worst-case execution time analysis
  - Program path analysis
  - Static analysis with value/cache/pipeline characterization
- Dynamic priority systems
  - EDF is optimal for certain cases, and the least utilization bound is 100%.
  - Demand bound analysis can be used to analyze the timing satisfaction.
- Static priority systems
  - Rate-monotonic is optimal for certain cases, and the least utilization bound is 69.3%.
  - Response time analysis can be used to analyze the worst-case response time.
- Resource sharing
  - Aperiodic tasks can be served by using resource reservation servers such that periodic tasks can still meet their timing constraints.
  - Priority inversion could kill the system and has to be handled.

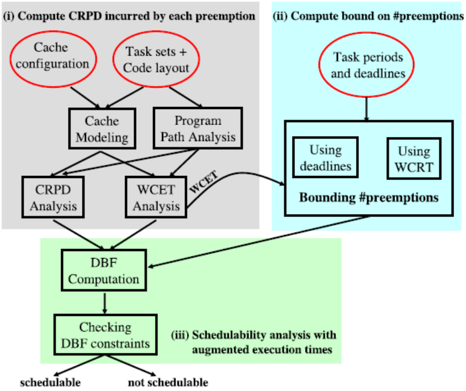
Things You May Already Forget

- Real-Time Calculus
  - More general mode of real-time tasks to bound the constraints of event arrivals in the interval domain.
  - Delay and buffer analysis can be done by finding the maximal horizontal and vertical distances between the service curve and the arrival curve.
- Partitioned scheduling
  - With resource augmentation, largest-task-first strategy can be very useful for identical multiprocessor systems.
- Semi-partitioned scheduling
  - With relaxation to execute a task on multiple processors, semi-partition can achieve better schedulability bounds.
- Global scheduling
  - Global EDF and RM are not helpful without resource augmentation.
  - The schedulability test can be done by considering the worst-case interference.

In the First Week....

- “Real time” is performance engineering/tuning.
  - Timeliness is more important in real-time systems.
- Real-time computing is equivalent to fast computing.
  - Real-time computing means predictable and reliable computing.
- There is no science in real-time system design.
  - What’s your opinion now?
- Advances in supercomputing hardware will take care of real-time requirements.
  - Buying a “faster” processor may result in timeliness violation.
- It is not meaningful to talk about guaranteeing real-time performance when things can fail.
  - Though hardwares may fail, the logic components, such as operating systems, should be solid when hardwares are still well functional.

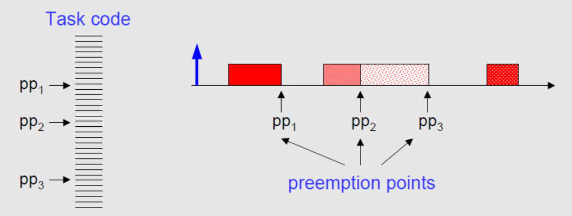
Things not Covered: Cache-Related Preemption Delay (CRPD)



- Preemption is not free
  - Cache is polluted
  - WCET analysis has to handle
  - Number of preemption has to be bounded to analyze worst-case response time
- Cache configuration matters

Lei Ju, Samarjit Chakraborty, Abhik Roychoudhury: Accounting for cache-related preemption delay in dynamic priority schedulability analysis. DATE 2007: 1623-1628

Things not Covered: Limited Preemption



- Preemption is only allowed at the preemption points
- Candidates for preemption points are important to reduce the cache-related preemption delay
- The non-preemptive segment might reduce the schedulability

Marko Bertogna, Giorgio C. Buttazzo, Mauro Marinoni, Gang Yao, Francesco Esposito, Marco Caccamo: Preemption Points Placement for Sporadic Task Sets. ECRTS 2010: 251-260

Sanjoy K. Baruah: The Limited-Preemption Uniprocessor Scheduling of Sporadic Task Systems. ECRTS 2005: 137-144

Things not Covered: Multiple Criticality Scheduling

Given a job J_i, with

- arrival time a_i
- relative deadline D_i
- criticality level L_i
  - Defense avionics: 2 (3?) criticalities, says safety-critical; mission-critical; non-critical
  - Civilian aviation (DO-178B): 5 criticalities, says catastrophic; hazardous; major; minor; no effect
  - Automotive systems (ISO 26262): 4 criticalities
- WCET function $C_i(1), C_i(2), \ldots$

Sanjoy K. Baruah, Haohan Li, Leen Stougie: Mixed-criticality Scheduling: Improved Resource-augmentation Results. CATA 2010: 217-223

Things not Covered: Multiple Criticality Scheduling

Given a job J_i, with

- arrival time a_i
- relative deadline D_i
- criticality level L_i
  - Defense avionics: 2 (3?) criticalities, says safety-critical; mission-critical; non-critical
  - Civilian aviation (DO-178B): 5 criticalities, says catastrophic; hazardous; major; minor; no effect
  - Automotive systems (ISO 26262): 4 criticalities
- WCET function $C_i(1), C_i(2), \ldots$

The MIXED-CRIT SCHEDULING PROBLEM: Given an instance $\{J_1, J_2, \ldots, J_n\}$ of mixed-criticality jobs, determine an appropriate scheduling strategy

Sanjoy K. Baruah, Haohan Li, Leen Stougie: Mixed-criticality Scheduling: Improved Resource-augmentation Results. CATA 2010: 217-223

Things not Covered: Soft Real-Time Guarantees

- hard real-time: No deadline is missed.
- soft real-time: Deadline tardiness is bounded, or the worst-case response time is bounded
  - There are many others
  - This one was proposed by Jim Anderson at UNC
  - You may use other definitions, but they are somehow difficult to be analyzed
  - Tasks still must execute sequentially, i.e., a tardy job of task τ_i cannot execute in parallel with the next job of τ_i.
  - A wide variety of global scheduling algorithms are capable of ensuring bounded tardiness with no utilization loss by using Global EDF.
  - Only negative results are known so far for Global fixed-priority scheduling.

Things not Covered: Global Scheduling on Uniform Processors

Why do we have to consider uniform processors?

- energy reduction
- different voltage islands

The details are omitted here.

Sanjoy K. Baruah: An Improved Global EDF Schedulability Test for Uniform Multiprocessors. IEEE Real-Time and Embedded Technology and Applications Symposium 2010: 184-192

Things not Covered: Cyber-Physical Systems

Edward A. Lee:

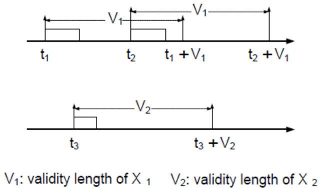
- Embedded software is software integrated with physical processes.
- The technical problem is managing time and concurrency in computational systems.
- Definition: Cyber-Physical (cy-phy) Systems (CPS) are integrations of computation with physical processes [Edward Lee, 2006].

CPS are usually

- real-time systems to react in time
- dynamic as the situation changes over time
- required joint optimization with “scheduling”, “control”, and “communication”.

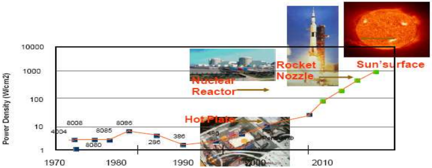
Things not Covered: Real-Time Database

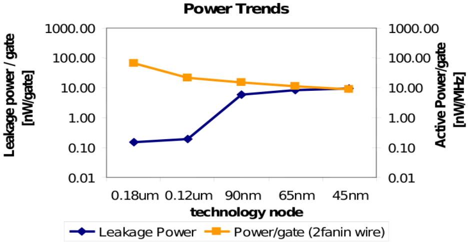
- A validity interval is associated with a data value.
- A data value is fresh within the validity interval.



- Half-Half [Ramamritham 93]: Period (T_i) and relative deadline (D_i) of an update transaction τ_i are each set to be one-half of the data validity length (V_i).
- More-Less [Burns & Davis 96, Xiong & Ramamritham 99]
  - Period + Relative Deadline $\leq$ Validity Length
  - Computation Time $\leq$ Relative Deadline $\leq$ Period
- Deferrable Scheduling with Fixed Priority (DS-FP) [Xiong, Han & Lam 05]
  - Adopts the sporadic task model and defers the sampling time as late as possible.

Things partially Covered: Energy and Power Issues





- Rapid increase of power consumption
- Slow increasing of the battery capacity
- Increasing cost of energy

- Dynamic power management (DPM)
- Micro-architecture technique
  - Adaptive architecture
  - Cache management
- Dynamic voltage scaling (DVS)
  - Supply voltage scaling
  - Threshold voltage scaling

Even More

- Resource reservation servers with worst-case performance guarantees
- Multiprocessor scheduling for general task models with resource augmentation guarantees
- Resource sharing protocols in multicore systems with worst-case guarantees
- Memory mapping and scheduling joint optimization
- Weakly hard real-time tasks, e.g., (m, k)-firm real-time tasks
- Hierarchical scheduling, e.g., FP+EDF+TDMA

Further Readings

- Open problems for Real-Time Scheduling Theory
  - Sanjoy K. Baruah, Kirk Pruhs: Open problems in real-time scheduling. J. Scheduling 13(6): 577-582 (2010)
- Summary for multiprocessor scheduling
  - Robert I. Davis, Alan Burns: A survey of hard real-time scheduling for multiprocessor systems. ACM Comput. Surv. 43(4): 35 (2011)

Questions?