Dynamic Scheduling Strategies in The ACE ORB

Christopher D. Gill
Washington University, St. Louis
cdgill@cs.wustl.edu

http://www.cs.wustl.edu/~cdgill/research/scheduling/gradpres.ps

28 May 1998
TAO's Static Scheduling Architecture

```
struct RT_Info {
    wc_exec_time_;  
cached_exec_time_;  
period_;          
importance_;      
dependencies_;    
};
```

1: CONSTRUCT CALL
   CHAINS OF RT_OPERATIONS
2: IDENTIFY THREADS

3: POPULATE
   RT_INFO
   REPOSITORY

4: ASSESS
   SCHEDULABILITY
5: ASSIGN OS THREAD
   PRIORITIES AND
   DISPATCH QUEUE
   ORDERING
   SUBPRIORITIES

6: SUPPLY
   PRIORITIES
   TO ORB

MODE 0
MODE 1
MODE 2
MODE 3

CURRENT
MODE
SELECTION

RUN-TIME
SCHEDULER
Priority/
Subpriority
Table Per
Mode
Limitations of Static Scheduling

- Assigning priority by rate limits achievable utilization
- Time cannot be reassigned if an operation is not called, or does not use its worst case computation time
- Goal: higher utilization
- Hypothesis: with dynamic scheduling techniques we can achieve this goal without undue overhead or instability of the schedule under load
Requirements for Dynamic Scheduling

- Achieve higher *utilization*
  - Schedule more of the unused time
- Preserve *stability* of the schedule under load
  - Isolate missed deadlines to operations that are not critical to the application
Operation Characteristics

- **Criticality** is an application defined significance of the operation missing its deadline
- **Period** is the time interval between arrivals of dispatch requests for the operation
- **Execution time** is the longest time used by one execution of the operation
- **Importance** is a weaker secondary indication of the operation’s significance
Generalizing Priority Assignment

- Priority assignment is a mapping from operation characteristics into urgency
- Each scheduling strategy provides a distinct mapping
  - Maximum Urgency First
  - Minimum Laxity First
  - Earliest Deadline First
  - Rate Monotonic Scheduling
  - Rate Monotonic—Dynamic
Maximum Urgency First Priority Assignment

- Static priority is assigned by criticality
- Dynamic subpriority is assigned according to minimum laxity
- Static subpriority is assigned by importance and finally by unique dispatch identifier
An RT_Info describes characteristics of one operation.

An RT_Operation wraps an RT_Info, and holds the set of all dispatches of that operation.

A Dispatch_Entry describes one dispatch of an operation.