

---

# Resource-bound process algebras for Schedulability and Performance Analysis of Real-Time and Embedded Systems

---

Insup Lee<sup>1</sup>, Oleg Sokolsky<sup>1</sup>, Anna Philippou<sup>2</sup>

<sup>1</sup> RTG (Real-Time Systems Group)  
Department of Computer and Information Science  
University of Pennsylvania  
Philadelphia, PA

<sup>2</sup> Department of Computer Science  
University of Cyprus  
Nicosia, CY

---

## Outline

---

- Real-Time and Embedded systems
- Resource-bound computation
- Resource-bound formalisms
  - ACSR (Algebra of communicating shared resources)
  - Schedulability Analysis Problem
  - PACSR (Probabilistic ACSR)
  - Schedulability analysis for soft real-time systems
  - Design framework for embedded systems
  - P<sup>2</sup>ACSR (Probabilistic ACSR with power consumption)
  - Scheduling synthesis and parametric schedulability analysis
  - ACSR-VP (ACSR with Value-Passing)
- Conclusions

## Real-time, Embedded Systems

---

- Difficulties
  - Increasing complexity
  - Decentralized
  - Safety critical
  - End-to-end timing constraints
  - Resource constrained
    - Non-functional: power, size, etc.
- Development of trustworthy (i.e., reliable, robust, safe, secure, etc.) embedded software

5/27/08

Korea University

3

## Properties of embedded systems

---

- Adherence to safety-critical properties
- Meeting timing constraints
- Satisfaction of resource constraints
- Confinement of resource accesses
- Supporting fault tolerance
- Domain specific requirements
  - Mobility
  - Software configuration

5/27/08

Korea University

4

## Real-time Behaviors

---

- Correctness and reliability of real-time systems depends on
  - Functional correctness
  - Temporal correctness
    - End-to-end temporal constraints
- Factors that affect temporal behavior are
  - Synchronization and communication
  - Resource limitations and availability/failures
  - Scheduling algorithms
  - Interaction with physical world
- An integrated framework to bridge the gap between concurrency theory and real-time scheduling

5/27/08

Korea University

5

## Scheduling Problems

---

- Priority Assignment Problem
- Schedulability Analysis Problem
  - Compositional analysis
  - Hierarchical system
- Soft timing/performance analysis (Probabilistic Performance Analysis)
- End-to-end Design Problem
  - Parametric Analysis
  - End-to-end constraints, intermediate timing constraints
  - Execution Synchronization Problem
  - Start-time Assignment Problem with Inter-job Temporal Constraints
- Fault tolerance: dealing with failures, overloads

5/27/08

Korea University

6

## Scheduling Factors

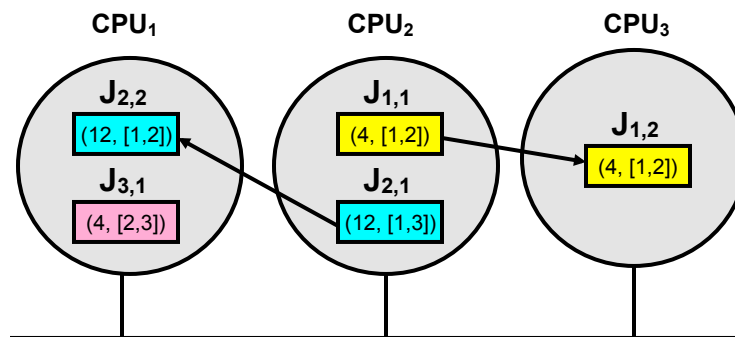
- Static priority vs dynamic priority
  - Cyclic executive, RM (Rate Monotonic)
  - EDF (Earliest Deadline First)
- Priority inversion problem
- Independent tasks vs. dependent tasks
- Single processor vs. multiple processors
- Communication delays
- Uncertainty in execution times
- Resource use tradeoffs
- End-to-end timing requirements

5/27/08

Korea University

7

## Example: Simple Scheduling Problem



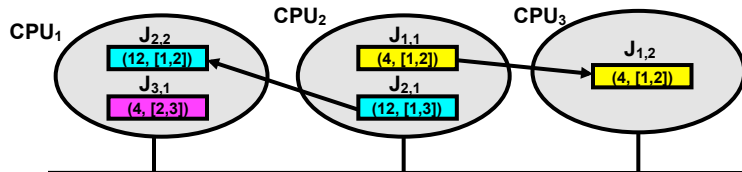
- ( period, [  $e^-$ ,  $e^+$  ] ), where  $e^-$  and  $e^+$  are the lower and upper bound of execution time, respectively.
- Goal is to find the priority of each job so that jobs are schedulable
- Considering only worst case leads to scheduling anomaly

5/27/08

Korea University

8

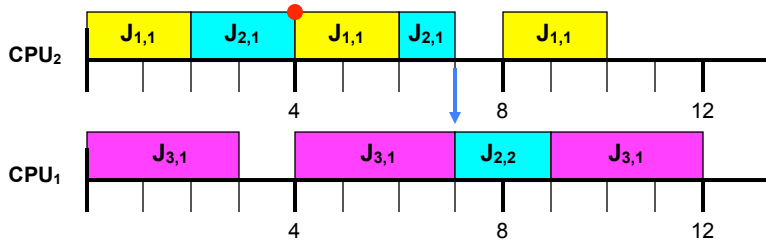
## Example (2)



Let  $J_{1,1} > J_{2,1}$  and  $J_{2,2} > J_{3,1}$

Consider worst case execution time for all jobs, i.e.,

Execution time  $E_{1,1} = 2, E_{2,1} = 3, E_{2,2} = 2, E_{3,1} = 3$

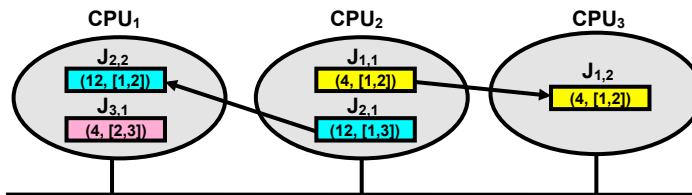


5/27/08

Korea University

9

## Example (3)



With same priorities,  $J_{1,1} > J_{2,1}$  and  $J_{2,2} > J_{3,1}$

Let execution time  $E_{1,1} = 1, E_{2,1} = 1, E_{2,2} = 2, E_{3,1} = 3$



So with the priority assignment of  $J_{1,1} > J_{2,1}$  and  $J_{2,2} > J_{3,1}$ , jobs cannot be scheduled and scheduling problems are in general NP-hard

5/27/08

Korea University

10

## End-to-end Design Problem

- Given a task set with end-to-end constraints on inputs and outputs
  - Freshness from input  $X$  to output  $Y$  ( $F(Y|X)$ ) constraints: bound time from input  $X$  to output  $Y$
  - Correlation between input  $X_1$  and  $X_2$  ( $C(Y|X_1, X_2)$ ) constraints: max time-skew between inputs to output
  - Separation between output  $Y$  ( $u(Y)$  and  $l(Y)$ ) constraints: separation between consecutive values on a single output  $Y$
- Derive scheduling for every task
  - Periods, offsets, deadlines
  - priorities
- Meet the end-to-end requirements
- Subject to
  - Resource limitations, e.g., memory, power, weight, bandwidth

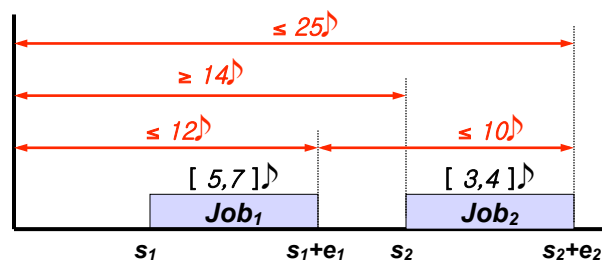
5/27/08

Korea University

11

## Example: Start-time Problem

Start-time Assignment Problem with Inter-job Temporal Constraints



**Goal** is to statically determine the range of start times for each job so that jobs are *schedulable* and all *inter-job temporal constraints* are satisfied.

5/27/08

Korea University

12

## Example: power-aware RT scheduling

- Dynamic Voltage Scaling allows tradeoffs between performance and power consumption
- Problem is how to minimize power consumption while meeting timing constraints.
- Example: three tasks with probabilistic execution time distribution

Task	Worst-case execution time	Period
1	3	8
2	3	10
3	2	14

5/27/08

Korea University

13

## Our approach and objectives

- Design formalisms for real-time and embedded systems
  - Resource-bound real-time process algebras
  - Executable specifications
  - Logic for specifying properties
- Design analysis techniques
  - Automated verification techniques
  - Parameterized end-to-end schedulability analysis
- Toolset implementation

5/27/08

Korea University

14

## Resource-bound computation

---

- Computational systems are always constrained in their behaviors
- Resources capture physical constraints
- Resources should be supported as a first-class notion in modeling and analysis
- Resource-bound computation is a general framework of wide applicability

## Resources

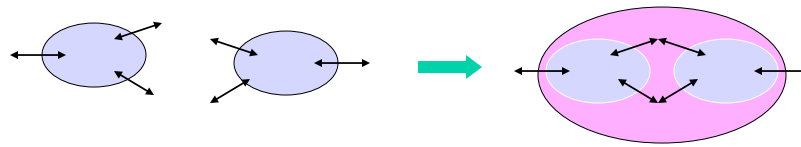
---

- **Resources** capture constraints on executions
- Resources can be
  - Serially reusable:
    - processors, memory, communication channels
  - Consumable
    - power
- Resource capacities
  - Single-capacity resources
  - Multiple-capacity resources
  - Time-sliced, etc.



## Process Algebras

- Process algebras are abstract and compositional methodologies for concurrent-system specification and analysis.
- "Design methodology which systematically allows to build complex systems from smaller ones" [Milner]



5/27/08

Korea University

17

## Process Algebras

- A process algebra consists of
  - a set of operators and syntactic rules for constructing processes
  - a semantic mapping which assigns meaning or interpretation to every process
  - a notion of equivalence or partial order between processes
  - a set of algebraic laws that allow syntactic manipulation of processes.
- Ancestors
  - CCS, CSP, ACP,...
  - focus on communication and concurrency

5/27/08

Korea University

18

## Advantages of Process Algebra

---

A large system can be broken into simpler subsystems and then proved correct in a modular fashion.

- 1 A hiding or restriction operator allows one to abstract away unnecessary details.
- 2 Equality for the process algebra is also a congruence relation; and thus, allows the substitution of one component with another equal component in large systems.

---

# ACSR

---

# ACSR

---

- **ACSR (Algebra of Communicating Shared Resource)**
  - A real-time process algebra which features discrete time, resources, and priorities
  - Timeouts, interrupts, and exception handling
  - Two types of actions:
    - Instantaneous events
    - Timed actions

# Events

---

- **Events** represent non-time consuming activities

- events are instantaneous: `crash`



- point-to-point synchronization



## Events

- **Events**

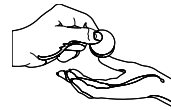
- have priorities:  $(\text{job}, 10^{10})$



- have input and output capabilities

or  $(e, p_1) \quad (\bar{e}, p_2)$

$(e?, p_1) \quad (e!, p_2)$



## Actions

- **Actions** represent activities that
  - take time
  - require access to resources
  - each resource usage has priority of access

$$A = \{(r_1, p_1), (r_2, p_2)\}$$

- each resource can be used at most once
- resources of action  $A$ :  $\rho(A)$
- idling action:  $\emptyset$

- Examples:

$\{(cpu, 2)\}, \{(cpu_1, 3), (cpu_2, 4)\},$   
 $\{(semaphore, 5)\}$

## Syntax for ACSR processes

- Process terms
  - Process names
- $$\begin{array}{l} P ::= \text{NIL} \\ | A : P \\ (a, n).P \\ P + P \\ P \parallel P \\ P \Delta_i^a(Q, R, S) \\ [P]_I \\ P \setminus F \\ b \rightarrow P \\ C \end{array}$$
- def*  
 $C = P$

5/27/08

Korea University

25

## Constant and Nil

*def*  
 $C = P$

C is a constant that represents the process algebra expression P

$P = \text{NIL}$

P does nothing

5/27/08

Korea University

26

## Prefix Operators

$P = A:Q$

P performs timed action A and then behaves as Q

$P = (a,n).Q$

P performs event (a,n) and then behaves as Q

### EXAMPLE

$\text{Operator} \stackrel{\text{def}}{=} (\text{ring},1).(\text{pickup},1).\text{Talk}$   
 $\text{Talk} = \{(\text{phone},2)\} : (\text{hangup},1).\text{Operator}$



5/27/08

Korea University

27

## Choice

$P = Q+R$

P can choose nondeterministically to behave like Q or R

### EXAMPLE



$\text{CAR} \stackrel{\text{def}}{=} (\text{golef},1).\text{CAR}'$   
 $+ (\text{goright},1).\text{CAR}''$

5/27/08

Korea University

28

## Parallel Composition

$P = Q \parallel R$

$P$  is composed by  $Q$  and  $R$  that may synchronize on events and must synchronize on timed actions

### EXAMPLE

$Operator \stackrel{def}{=} (ring?,1). \{(phone,2)\}$   
 $:(hangup?,1).Operator$

$Caller \stackrel{def}{=} (ring!,2). \{(phone!,3)\}$   
 $:(hangup!,1).Caller$

$Converse \stackrel{def}{=} Operator \parallel Caller$



5/27/08

Korea University

29

## Scope

$P = Q \Delta_t^a (R, S, T)$

$Q$  may execute for at most  $t$  time units. If message  $a$  is produced, control is delegated to  $R$ , else control is delegated to  $S$ . At any time  $T$  may interrupt.

### EXAMPLE

$Runner \stackrel{def}{=} Run \Delta_{10}^{finish} (GoForCoffee,$   
 $GoToWork,$   
 $BeepedToWork)$

$Run \stackrel{def}{=} \{(run,1)\} : Run + finish!.NIL$



5/27/08

Korea University

30

## Hiding/Restriction

$$P = [Q]_I$$

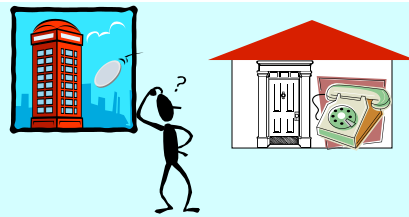
P behaves just as Q but resources in I are no longer visible to the environment

$$P = Q \setminus F$$

P behaves just as Q but labels in F are no longer visible to the environment

### EXAMPLE

Caller || PayPhone || [Home]<sub>phone</sub>



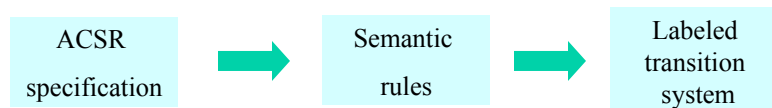
5/27/08

Korea University

31

## ACSR semantics

- Gives an unambiguous meaning to language expressions.
- Semantics is operational, given by a set of semantic rules.



- Example of a labeled transition system:

$$P_0 \xrightarrow{\emptyset} P_1 \xrightarrow{NC} P_2 \xrightarrow{\{gate, train\}} P_3 \xrightarrow{\{gate, train\}} P_4 \xrightarrow{IC} \dots$$

5/27/08

Korea University

32



## ACSR semantics

---

- Two-level semantics:

- A collection of inference rules gives the **unprioritized** transition relation

$$P \xrightarrow{\alpha} P'$$

- A **preemption** relation on actions and events disables some of the transitions, giving a **prioritized** transition relation

$$P \xrightarrow{\alpha}_{\pi} P'$$

## Unprioritized transition relation

---

- Prefix operators

$$\mathbf{ActT} \frac{-}{A: P \xrightarrow{A} P}$$

$$\mathbf{ActI} \frac{-}{(a, p): P \xrightarrow{(a, p)} P}$$

- Choice

$$\mathbf{ChoiceL} \frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'}$$

- Parallel

$$\mathbf{ParIL} \frac{P \xrightarrow{(a, p)} P'}{P \parallel Q \xrightarrow{(a, p)} P' \parallel Q}$$

## Unprioritized transition relation (II)

- Resource-constrained execution

$$\text{ParT} \frac{P \xrightarrow{A_1} P' \quad Q \xrightarrow{A_2} Q'}{P \parallel Q \xrightarrow{A_1 \cup A_2} P' \parallel Q'} \quad \rho(A_1) \cap \rho(A_2) = \emptyset$$

- Priority-based communication

$$\text{ParCom} \frac{P \xrightarrow{(a?, p_1)} P' \quad Q \xrightarrow{(a!, p_2)} Q'}{P \parallel Q \xrightarrow{(a, p_1 + p_2)} P' \parallel Q'}$$

- Resource closure

$$\text{CloseT} \frac{P \xrightarrow{A_1} P'}{[P]_I \xrightarrow{A_1 \cup A_2} [P']_I} \quad A_2 = \{(r, 0) \mid r \in I - A_1\}$$

5/27/08

Korea University

35

## Examples

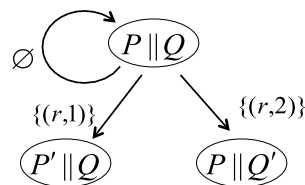
- Resource conflict

$$P = \{(r, 1)\} : P' \quad Q = \{(r, 2)\} : Q' \quad P \parallel Q \sim \text{NIL}$$

- Processes must provide for preemption

$$P = \{(r, 1)\} : P' + \emptyset : P \quad Q = \{(r, 2)\} : Q' + \emptyset : Q$$

- Unprioritized transitions:



5/27/08

Korea University

36

## Unprioritized transition relation (III)

$$\text{ScopeCT} \frac{P \xrightarrow{A} P'}{P\Delta_t^a(Q, R, S) \xrightarrow{A} P'\Delta_{t-1}^a(Q, R, S)} \quad (t > 0)$$

$$\text{ScopeCI} \frac{P \xrightarrow{e} P'}{P\Delta_t^a(Q, R, S) \xrightarrow{e} P'\Delta_t^a(Q, R, S)} \quad (l(e) \neq a, t > 0)$$

$$\text{ScopeE} \frac{P \xrightarrow{(a,n)} P'}{P\Delta_t^a(Q, R, S) \xrightarrow{(\tau,n)} Q} \quad (t > 0)$$

$$\text{ScopeT} \frac{R \xrightarrow{\alpha} R'}{P\Delta_t^a(Q, R, S) \xrightarrow{\alpha} R'} \quad (t = 0)$$

$$\text{ScopeI} \frac{S \xrightarrow{\alpha} S'}{P\Delta_t^a(Q, R, S) \xrightarrow{\alpha} S'} \quad (t > 0)$$

5/27/08

Korea University

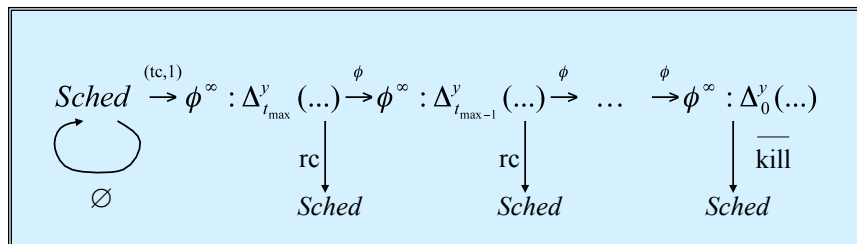
37

## Example

- A Scheduler

$$\text{Sched} = \phi : \text{Sched}$$

$$+ (tc, 1). \phi^\infty \Delta_{\max}^y (\overline{NIL}, \overline{kill} . \text{Sched}, rc . \text{Sched})$$



5/27/08

Korea University

38

## Preemption relation

- To take priorities into account in the semantics we define the relation  $\alpha$  is preempted by  $\beta$ :  $\alpha < \beta$
- An action  $\beta$  preempts an action  $\alpha$  iff
  - no lower priorities:  $\forall r \in \rho(\alpha), \pi_r(\alpha) \leq \pi_r(\beta)$
  - some higher priorities:  $\exists r \in \rho(\beta), \pi_r(\alpha) < \pi_r(\beta)$
  - it contains fewer resources  $\rho(\beta) \subseteq \rho(\alpha)$
 e.g.  $\{(r_1, 3), (r_2, 5)\} < \{(r_1, 7), (r_2, 5)\}$
- An event preempts another event iff
  - same label, higher priority e.g.  $(a!, 1) < (a!, 3)$
- An event preempts an action iff
  - $\tau$  with non-zero priority preempts all actions e.g.  $\{(r, 4)\} < (\tau, 1)$

5/27/08

Korea University

39

## Prioritized transition relation

- We define  $P \xrightarrow{\alpha} \pi P'$
- when
  - there is an unprioritized transition  $P \xrightarrow{\alpha} P'$
  - there is no  $P \xrightarrow{\beta} P''$  such that  $\alpha < \beta$
- Compositional

5/27/08

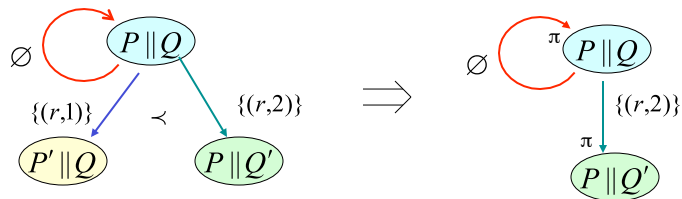
Korea University

40

## Example

- Unprioritized and prioritized transitions:

$$P = \{(r,1)\} : P' + \emptyset : P \quad Q = \{(r,2)\} : Q' + \emptyset : Q$$



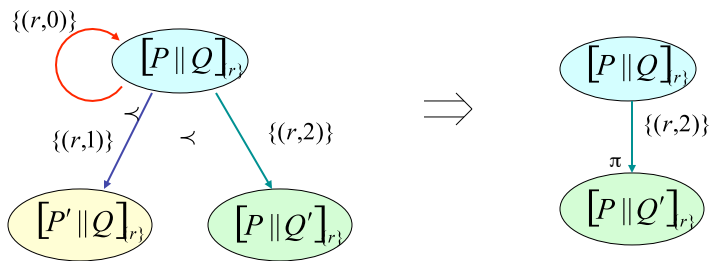
5/27/08

Korea University

41

## Example (cont.)

- Resource closure enforces progress



5/27/08

Korea University

42

## Compositionality of preemption relation

- Given

$$P_1 = (a,2).S_1 + (b,1).S_2$$

$$P_2 = (a,2).S_1$$

$$Q_1 = (\bar{a},3).T_1 + (\bar{b},5).T_2$$

$$Q_2 = (\bar{a},3).T_1 + (\bar{b},2).T_2$$

$$R_1 = (b,2).S_1 + (b,1).S_2$$

$$R_2 = (b,2).S_1$$

- Given  $P_1$  and  $P_2$ , can they be treated as equivalent?  
That is, for all  $Q$ ,  $P_1 \parallel Q = P_2 \parallel Q$ ?
- How about  $R_1$  and  $R_2$ ?

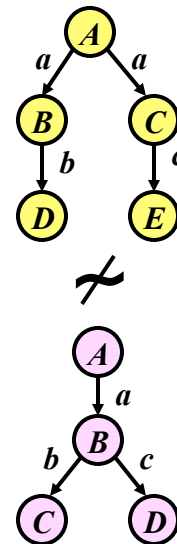
5/27/08

Korea University

43

## Bisimulation

- Observational equivalence** is based on the idea that two equivalent systems exhibit the same behavior at their interfaces with the environment.
- This requirement was captured formally through the notion of **bisimulation**, a binary relation on the states of systems.
- Two states are **bisimilar** if for each single computational step of the one there exists an appropriate matching (multiple) step of the other, leading to bisimilar states.



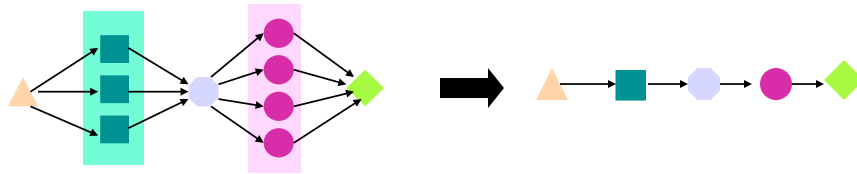
5/27/08

Korea University

44

## Prioritized strong equivalence

- An equivalence relation is congruence when it is preserved by all the operators of the language.
- This implies that replacement of equivalent components in any complex system leads to equivalent behavior.



- Strong bisimulation  $\sim_{\pi}$  over  $P \xrightarrow{\alpha} P'$  is a congruence relation with respect to the ACSR operators.

5/27/08

Korea University

45

## Equational Laws

- Equational laws are a set of axioms on the syntactic level of the language that characterize the equivalence relation.
- They may be used for manipulating complex systems at the level of their syntactic (ACSR) description.
- There is a set of laws that is complete for finite state ACSR processes:

$$\begin{array}{ll}
 P + NIL = P & P + P = P \\
 P + Q = Q + P & (P \parallel Q) \parallel R = P \parallel (Q \parallel R) \\
 \dots &
 \end{array}$$

5/27/08

Korea University

46

## Equational Laws

---

- ACSR-specific laws for scope and resource closure:

$$\begin{aligned}
 A: P\Delta_t^a(Q, R, S) &= A: (P\Delta_{t-1}^a(Q, R, S)) \dagger S && \text{if } t > 0 \\
 e.P\Delta_t^a(Q, R, S) &= e. (P\Delta_t^a(Q, R, S)) \dagger S && \text{if } t > 0 \wedge \overline{l(e)} \neq a \\
 e.P\Delta_t^a(Q, R, S) &= (\tau, \pi(e))Q \dagger S && \text{if } t > 0 \wedge \overline{l(e)} = a \\
 P\Delta_0^a(Q, R, S) &= R \\
 \left[ \begin{array}{l} A_1 : P \\ e.P \end{array} \right] &= (A_1 \cup A_2) : P && A_2 = \{ (r, 0) \mid r \in I - \rho(A_1) \} \\
 &= e.[P]
 \end{aligned}$$

5/27/08

Korea University

47

## Laws (1)

---

$$\begin{aligned}
 \text{Choice(1)} \quad & P + \text{NIL} = P \\
 \text{Choice(2)} \quad & P + P = P \\
 \text{Choice(3)} \quad & P + Q = Q + P \\
 \text{Choice(4)} \quad & (P + Q) + R = P + (Q + R) \\
 \text{Choice(5)} \quad & \alpha P + \beta Q = \beta Q \quad \text{if } \alpha \prec \beta \\
 \text{Par(1)} \quad & P \parallel Q = Q \parallel P \\
 \text{Par(2)} \quad & (P \parallel Q) \parallel R = P \parallel (Q \parallel R) \\
 \text{Par(3)} \quad & \left( \sum_{i \in I} A_i : P_i + \sum_{j \in J} e_j \cdot Q_j \right) \parallel \left( \sum_{k \in K} B_k : R_k + \sum_{l \in L} f_l : S_l \right) \\
 &= \left[ \begin{array}{l} \sum_{\substack{i \in I, k \in K \\ \rho(A_i) \cap \rho(B_k) = \emptyset}} (A_i : B_k) : (P_i \parallel R_k) \\ + \sum_{j \in J} e_j \cdot (Q_j \parallel (\sum_{k \in K} B_k : R_k + \sum_{l \in L} f_l : S_l)) \\ + \sum_{l \in L} f_l \cdot ((\sum_{i \in I} A_i : P_i + \sum_{j \in J} e_j \cdot Q_j) \parallel S_l) \\ \sum_{\substack{j \in J, l \in L \\ l(e_j) = l(f_l)}} (\tau, \pi(e_j) + \pi(f_l)) \cdot (Q_j \parallel S_l) \end{array} \right]
 \end{aligned}$$

5/27/08

Korea University

48



## Laws (2)

---

- Scope(1)  $A : P\Delta_t^b(Q, R, S) = A : (P\Delta_{t-1}^b(Q, R, S)) + S$  if  $t > 0$
- Scope(2)  $e.P\Delta_t^b(Q, R, S) = e.(P\Delta_{t-1}^b(Q, R, S)) + S$  if  $t > 0 \wedge \overline{l(e)} \neq b$
- Scope(3)  $e.P\Delta_t^b(Q, R, S) = (\tau, \pi(e)).\underline{Q} + S$  if  $t > 0 \wedge \overline{l(e)} = b$
- Scope(4)  $P\Delta_0^b(Q, R, S) = R$
- Scope(5)  $(P_1 + P_2)\Delta_t^b(Q, R, S) = P_1\Delta_t^b(Q, R, S) + P_2\Delta_t^b(Q, R, S)$
- Scope(6)  $NIL\Delta_t^b(Q, R, S) = S$  if  $t > 0$
- Res(1)  $NIL \setminus F = NIL$
- Res(2)  $(P + Q) \setminus F = (P \setminus F) + (Q \setminus F)$
- Res(3)  $(A : P) \setminus F = A : (P \setminus F)$
- Res(4)  $((a, n).P) \setminus F = (a, n).(P \setminus F)$  if  $a, \bar{a} \notin F$
- Res(5)  $((a, n).P) \setminus F = NIL$  if  $a, \bar{a} \in F$
- Res(6)  $P \setminus E \setminus F = P \setminus E \cup F$
- Res(7)  $P \setminus \emptyset = P$

5/27/08

Korea University

49

## Laws (3)

---

- Close(1)  $[NIL]_I = NIL$
- Close(2)  $[P + Q]_I = [P]_I + [Q]_I$
- Close(3)  $[A_1 : P]_I = (A_1 \cup A_2) : [P]_I$  where  $A_2 = \{(r, 0) \mid r \in I - \rho(A_1)\}$
- Close(4)  $[e.P]_I = e.[P]_I$
- Close(5)  $[[P]_I]_J = [P]_{I \cup J}$
- Close(6)  $[P]_\emptyset = P$
- Close(7)  $[P \setminus E]_I = [P]_I \setminus E$
- Rec(1)  $rec X.P = P[rec X.P / X]$
- Rec(2) If  $P = Q[P / X]$  and  $X$  is guarded in  $Q$  then  $P = rec X.Q$
- Rec(3)  $rec X.(P + \sum_{i \in I} [X \setminus E_i]_{U_i}) = rec X.(\sum_{j \in J} [P \setminus E_j]_{U_j})$
- where  $E_j = \bigcup_{i \in I} E_i, U_j = \bigcup_{i \in I} U_i, I$  is finite and  $X$  is guarded in  $P$

5/27/08

Korea University

50

## Soundness of the laws

---

- **Theorem:**

if  $P=Q$  then  $P \sim_{\pi} Q$

- Proof approach:

- Construct the set of prioritized derivations for each P
- Prove that if  $P=Q$ , then the sets of derivations are the same

## Completeness of the laws

---

- **Theorem:**

if P and Q are finite-state processes and  $P \sim_{\pi} Q$   
then  $P=Q$

---

## Schedulability Analysis

---

---

## Schedulability Analysis

---

- Can all real-time tasks meet their deadlines?
- Factors include
  - Delay caused by synchronization between tasks
  - Delay caused by precedence between tasks
  - Delay caused by resource constraints
  - Scheduling disciplines and synchronization protocols

## Outline

---

- ACSR-VP: ACSR with value-passing and dynamic priorities
- Specifying real-time systems using ACSR-VP
  - Specifying task models
  - Specifying scheduling disciplines
- Analyzing real-time systems using bisimulation
  - Specification correctness
  - Schedulability analysis
- Schedulability analysis using VERSA (ACSR Toolkit)

5/27/08

Korea University

55

## ACSR (Algebra of Communicating Shared Resources)

---

- A timed process algebra based on CCS with notions of time, resources and priorities
- Discrete time and dense time
- Static priorities
- Actions: Instantaneous Events + Timed Actions
  - Timed action: a set of (resource, priority) pairs  
 $\{(cpu, 4), (data, 3)\}, \{(cpu_1, 2), (cpu_2, 3)\}, \emptyset$
  - Instantaneous event: (event, priority) pair  
 $(signal, 2), (chan, 2), (\tau, 3)$
- Real-time operators for timeout, interrupt, exception
- Graphical specification language (GCSR)
- Toolkit (VERSA)
- No value passing communication, no variables for priorities

5/27/08

Korea University

56

## ACSR-VP (ACSR with Value Passing)

- Extends ACSR with variables and value passing communications
- Values can be specified using expressions
  - Timed Actions:  
 $\{(cpu, x), (data, y + 1)\}$
  - Instantaneous events:  
 $(signal!8, x)$  - output  
 $(chan?y, 2)$  - input
- Dynamic priorities
- Exchange priority information without global variables

5/27/08

Korea University

57

## ACSR-VP Syntax

$P ::=$	$NIL$	process that does nothing
	$A:P$	timed action prefix
	$e.P$	instantaneous event prefix
	$be \rightarrow P$	conditional process ( $be$ : boolean expression)
	$P_1 + P_2$	choice
	$P_1 \parallel P_2$	parallel composition
	$[P]_I$	resource close
	$P \setminus F$	event restriction
	$P \setminus I$	resource hiding
	$C(\underline{x})$	process name defined to be a process $C(\underline{x}) = P$

5/27/08

Korea University

58

## ACSR-VP Example

### Preemptable and Non-preemptable Jobs

- Both jobs execute  $c$  time units on  $cpu$  with priority  $\pi$
- Non-preemptable job: once it acquires  $cpu$ , it executes to completion

$$\begin{aligned} \text{Job}_1 &\stackrel{\text{def}}{=} \emptyset : \text{Job}_1 + \text{Exec}_1(0) \\ \text{Exec}_1(s) &\stackrel{\text{def}}{=} (s < c) \rightarrow \{(cpu, \pi)\} : \text{Exec}_1(s+1) \end{aligned}$$

- Preemptable job: its execution can be preempted by actions on  $cpu$  of other jobs with higher priorities

$$\begin{aligned} \text{Job}_2 &\stackrel{\text{def}}{=} \emptyset : \text{Job}_2 + \text{Exec}_2(0) \\ \text{Exec}_2(s) &\stackrel{\text{def}}{=} (s < c) \rightarrow \{(cpu, \pi)\} : \text{Exec}_2(s+1) \\ &\quad + \emptyset : \text{Exec}_2(s) \end{aligned}$$

## Unprioritized Operational Semantics

$$\text{Act} \quad A : P \xrightarrow{A} P$$

$$\text{ActI1} \quad (l? \langle \underline{x} \rangle, ve). P \xrightarrow{(l? \langle n \rangle, [ve])} P[\underline{n} / \underline{x}]$$

$$\text{ActI2} \quad (l! \langle \underline{ve}_2 \rangle, ve_1). P \xrightarrow{(l! [\underline{ve}_2], [ve_1])} P$$

$$\text{ActI3} \quad (\tau, ve). P \xrightarrow{(\tau, [ve])} P$$

$$\text{ParT} \quad \frac{P \xrightarrow{A_1} P', Q \xrightarrow{A_2} Q'}{P \parallel Q \xrightarrow{A_1 \cup A_2} P' \parallel Q'} \quad (\rho(A_1) \cap \rho(A_2) = \emptyset)$$

$$\text{ParC2} \quad \frac{P \xrightarrow{(l! \langle \underline{k} \rangle, m)} P', Q \xrightarrow{(l? \langle \underline{k} \rangle, n)} Q'}{P \parallel Q \xrightarrow{(\tau, m+n)} P' \parallel Q'}$$

## Unprioritized Operational Semantics

$$\begin{array}{l}
 \text{CloseT} \quad \frac{P \xrightarrow{A_1} P'}{\boxed{P} \xrightarrow{A_1 \cup A_2} \boxed{P'}} \quad (A_2 = \{(r,0) \mid r \in I - \rho(A_1)\}) \\
 \text{CloseI} \quad \frac{P \xrightarrow{e} P'}{\boxed{P} \xrightarrow{e} \boxed{P'}} \\
 \text{HideT} \quad \frac{P \xrightarrow{A} P'}{P \setminus \setminus I \xrightarrow{A'} P \setminus \setminus I} \quad (\{(r,p) \in A \mid r \notin I\}) \\
 \text{HideI} \quad \frac{P \xrightarrow{e} P'}{P \setminus \setminus I \xrightarrow{e} P \setminus \setminus I}
 \end{array}$$

5/27/08

Korea University

61

## Preemption

A preemption relation is defined for two any actions  $\alpha$  and  $\beta$ , denoted  $\alpha \prec \beta$ , read  $\beta$  preempts  $\alpha$ .

Examples:

- $\{(r_1,2), (r_2,5)\} \prec \{(r_1,7), (r_2,5)\}$
- $\{(r_1,2), (r_2,5)\} \not\prec \{(r_1,7), (r_2,3)\}$
- $\{(r_1,2), (r_2,0)\} \prec \{(r_1,7)\}$
- $\{(r_1,2), (r_2,1)\} \not\prec \{(r_1,7)\}$
- $(a,2) \prec (a,5)$
- $(a,1) \not\prec (b,2)$
- $(\tau,1) \prec (\tau,2)$
- $\{(r_1,2), (r_2,5)\} \prec (\tau,2)$

5/27/08

Korea University

62

## Prioritized Operational Semantics

The operational semantics of ACSR-VP, the *prioritized transition relation*  $\xrightarrow{\alpha}_{\pi}$  is defined as follows:

$P \xrightarrow{\alpha}_{\pi} P'$  iff (1)  $P \xrightarrow{\alpha} P'$   
 (2) there is no  $P \xrightarrow{\beta} P''$  such that  $\alpha < \beta$

Example:  $P \stackrel{def}{=} \{(cpu,2)\} : P_1 + \{(cpu,3)\} : P_2$

- Unprioritized transition :  $\begin{cases} P \xrightarrow{\{(cpu,2)\}} P_1 \\ P \xrightarrow{\{(cpu,3)\}} P_2 \end{cases}$
- Prioritized transition :  $P \xrightarrow{\{(cpu,3)\}}_{\pi} P_2$

## Modeling a Real-Time System

- A real-time system consists of a set of tasks running in parallel under a specific scheduling discipline
- A task is a process composed of a sequence of jobs executed serially
- A task can be
  - Independent or dependent
  - Preemptable or non-preemptable
  - Periodic or aperiodic
- Possible timing constraints of a task are:

b	Starting time
c, d	Execution time and deadline
p	Period for periodic task
$p_1, p_2$	Minimum and maximum inter - arrival times for aperiodic task



## Specification of a real-Time System

A real-time system is specified by the process **RTS**:

$$\text{RTS} \stackrel{\text{def}}{=} [T_1 \parallel T_2 \parallel \dots \parallel T_n]$$

Tasks are specified by the processes  $T_i$ :

$$T_i \stackrel{\text{def}}{=} (\text{Job}_i \parallel \text{Activator}_i) \setminus \{start, end\}$$

- Process **Job<sub>i</sub>**: internal characteristics, e.g.:
  - resource requirements
  - synchronization
- Process **Activator<sub>i</sub>**: external timing attributes, e.g.,
  - periodic or aperiodic
  - period and deadline
- Events *start*, *end* are synchronization events:
  - *start*: activate jobs
  - *end*: mark deadlines of jobs - deadlock if unsuccessful

## Sample Activators

Activator 1. A periodic task with (b, d, p)

$$\begin{aligned} \text{Activator} &\stackrel{\text{def}}{=} \emptyset^b : \text{Activator}' \\ \text{Activator}' &\stackrel{\text{def}}{=} (start!, 1). \emptyset^d : (end!, 2). \\ &\quad \emptyset^{p-d} : \text{Activator}' \end{aligned}$$

Activator 2. An aperiodic task with (b, d, p<sub>1</sub>, p<sub>2</sub>)

$$\begin{aligned} \text{Activator} &\stackrel{\text{def}}{=} \emptyset^b : \text{Activator}' \\ \text{Activator}' &\stackrel{\text{def}}{=} (start!, 1). \emptyset^d : (end!, 2). \\ &\quad \emptyset^{p_1-d \dots p_2-d} : \text{Activator}' \end{aligned}$$

where

$$\begin{aligned} \emptyset^n &\stackrel{\text{def}}{=} \emptyset : \dots : \emptyset \quad (\text{idling for } n \text{ time units}) \\ \emptyset^{m..n} &\stackrel{\text{def}}{=} \emptyset^m + \emptyset^{m+1} + \dots + \emptyset^n \end{aligned}$$

## Sample Jobs

---

### Job 1

- preemptable, independent jobs running on *cpu* priority  $\pi$  and execution time  $c$ :

$$\begin{aligned} \text{Job} &\stackrel{\text{def}}{=} \emptyset : \text{Job} + (\text{start}?,1).\text{Exec}(0,0) \\ \text{Exec}(s,t) &\stackrel{\text{def}}{=} (s < c) \rightarrow (\{cpu, \pi\} : \text{Exec}(s+1, t+1) \\ &\quad + \emptyset : \text{Exec}(s, t+1)) \\ &\quad + (s = c) \rightarrow \text{Wait} \\ \text{Wait} &\stackrel{\text{def}}{=} \emptyset : \text{Wait} + (\text{end}?,1).\text{Job} \end{aligned}$$

- $s$  for accumulated execution time
- $t$  for the elapsed time
- **Job** can response to *end* event only when its current execution is finished

5/27/08

Korea University

67

## Sample Jobs

---

### Job 2

- nonpreemptable, independent jobs on multiprocessors  $cpu_1, \dots, cpu_k$  with priorities  $\pi_1, \dots, \pi_k$  and execution time  $c$ :

$$\begin{aligned} \text{Job} &\stackrel{\text{def}}{=} \emptyset : \text{Job} + (\text{start}?,1).\text{Exec} \\ \text{Exec} &\stackrel{\text{def}}{=} \sum_{1 \leq i \leq k} (\{cpu_i, \pi_i\}^c : \text{Wait}) \\ \text{Wait} &\stackrel{\text{def}}{=} \emptyset : \text{Wait} + (\text{end}?,1).\text{Job} \end{aligned}$$

- A job can be executed on any of the processors
- Once a processor is assigned to a job, the job executes on that processor until completion

5/27/08

Korea University

68

## Sample Jobs

---

### Job 3

- dependent jobs on processor  $cpu$  with priority  $\pi$  and execution time  $c$  a single preemptible critical section of length  $cs$  on resource  $data$  (with priority  $\pi'$ ) after at  $c'$  time units execution:

$Job$	$\stackrel{def}{=} \emptyset : Job + (start?,1).Exec(0,0)$
$Exec$	$\stackrel{def}{=} (s < c \wedge s \neq c') \rightarrow (\{cpu, \pi\} : Exec(s+1, t+1) + \emptyset : Exec(s, t+1))$ $+ (s = c') \rightarrow (\{p!, 0\}.CS(s, t) + \emptyset : Exec(s, t+1))$ $+ (s = c) \rightarrow Wait$
$Wait$	$\stackrel{def}{=} \emptyset : Wait + (end?,1).Job$
$CS(s, t)$	$\stackrel{def}{=} (s < c' + cs) \rightarrow (\{cpu, \pi'\} : CS(s+1, t+1) + \emptyset : CS(s, t+1))$ $+ (s = c' + cs) \rightarrow (v!, 0).Exec(s, t)$
$P$	$\stackrel{def}{=} (p?, 0).V + \emptyset : P$
$V$	$\stackrel{def}{=} (v?, 0).P + \emptyset : V$

- $P$  and  $V$  operations are modeled by the processes  $P$  and  $V$  with events  $(p?, 0)$  and  $(v?, 0)$
- When  $s$  equals  $c'$ ,  $Exec$  waits for  $(p?, 0)$  to enter the critical section  $CS(s, t)$

5/27/08

Korea University

69

## Scheduling Disciplines

---

### Earliest Deadline First

- Tasks  $T_i \stackrel{def}{=} Job_i + Activator_i$
  - Priority  $\pi_i \stackrel{def}{=} d_{max} - (d_i - t)$
- where  $d_{max} \stackrel{def}{=} (1 + \max\{d_1, \dots, d_n\})$

$EDFSys$	$\stackrel{def}{=} [T_1 \parallel T_2 \parallel \dots \parallel T_n]_{cpu}$
$T_i$	$\stackrel{def}{=} (Job_i \parallel Activator_i) \setminus \{start, end\}$
$Job_i$	$\stackrel{def}{=} \emptyset : Job_i + (start?, 1).Exec_i(0, 0)$
$Exec_i(s, t)$	$\stackrel{def}{=} (s < c_i) \rightarrow (\{cpu, d_{max} - (d_i - t)\} : Exec_i(s, t+1) + \emptyset : Exec_i(s, t+1))$ $+ (s = c) \rightarrow Wait$
$Wait_i$	$\stackrel{def}{=} \emptyset : Wait_i + (end?, 1).Job_i$
$Activator_i$	$\stackrel{def}{=} (start!, 1).\emptyset^{d_i} : (end!, 2).\emptyset^{p-d} : Activator_i$

5/27/08

Korea University

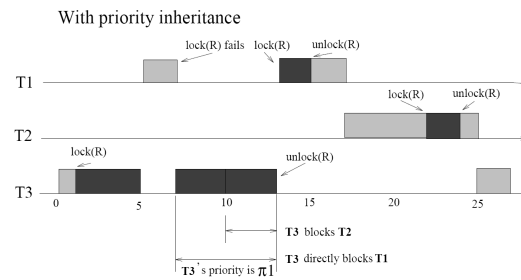
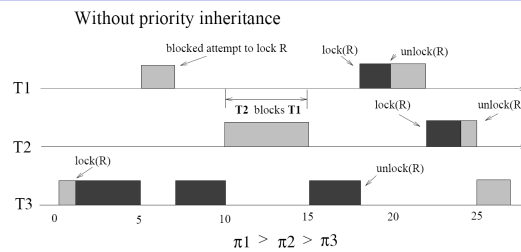
70

## Other Time-Driven Scheduling Disciplines

Deadline Monotonic	$\pi_i \stackrel{def}{=} d_{max} - d_i$
Shortest Remaining Time First	$\pi_i \stackrel{def}{=} c_{max} - (c_i - s)$
Least Laxity First	$\pi_i \stackrel{def}{=} d_{max} - (d_i - t) - (c_i - s)$

where  $c_{max} \stackrel{def}{=} (1 + \max\{c_1, \dots, c_n\})$

## The Priority Inversion Problem



## Task parameters

Resources :	<i>cpu</i>	processor		
Constants :	ready time	: $r_1 = 5$	$r_2 = 10$	$r_3 = 0$
	comp. time	: $c_1 = 6$	$c_2 = 8$	$c_3 = 13$
	deadline	: $d_1 = 30$	$d_2 = 30$	$d_3 = 30$
	start time of CS	: $cs_1 = 3$	$cs_2 = 5$	$cs_3 = 1$
	length of CS	: $c'_1 = 2$	$c'_2 = 2$	$c'_3 = 10$
	priority	: $\pi_1 = 3$	$\pi_2 = 2$	$\pi_3 = 1$
	max priority	: $\pi_{max} = 4$		

## Priority Inheritance Protocol

$T_i \stackrel{def}{=} \text{Job } 3 + \text{Activator } 1 + \text{Priority - Passing Events}$	
$PIPSys$	$\stackrel{def}{=} \left[ \left[ T_1 \parallel T_2 \parallel T_3 \parallel P \right] \{req, chan, p, v\}_{cpu} \right]$
$T_i$	$\stackrel{def}{=} (\text{Job}_i \parallel \text{Activator}_i) \{start, end\}$
$\text{Job}_i$	$\stackrel{def}{=} \emptyset : \text{Job}_i + (start?1).Exec.(0,0)$
$Exec_i(s)$	$\stackrel{def}{=} (s < c_i \wedge s \neq cs_i) \rightarrow (\{cpu, \pi_i\}) : Exec.(s+1) + \emptyset : Exec.(s)$ $+ (s = cs_i) \rightarrow ((req! \pi_i, \pi_i).Req_i(s) + \emptyset : Exec_i(s)) + (s = c_i) \rightarrow \text{Wait}$
$\text{Wait}_i$	$\stackrel{def}{=} \emptyset : \text{Wait}_i + (end?1).Job_i$
$Req_i(s)$	$\stackrel{def}{=} (p! \pi_i, \pi_i).CS_i(s, \pi_i) + \emptyset : Req_i(s)$
$CS_i(s, \pi)$	$\stackrel{def}{=} (s < c'_i + cs_i) \rightarrow (\{cpu, \pi\}) : CS_i(s+1, \pi) + (chan? new; 1).CS_i(s, new) + \emptyset : CS_i(s, \pi)$ $+ (s = c'_i + cs_i) \rightarrow (v?1).Exec_i(s)$
$\text{Activator}_i$	$\stackrel{def}{=} \emptyset^i : (start!1).\emptyset^{di} : (end!2).\emptyset^\infty$
$P$	$\stackrel{def}{=} (p?x1).V(x) + (req?x, \pi_{max}).(p?x, 1).V(x) + \emptyset : P$
$V(max)$	$\stackrel{def}{=} (v!1).P + \emptyset : V(max) + (req?x, 1).(x > max) \rightarrow (chan!x, 1).V(x) + (x \leq max) \rightarrow V(max)$

Parameters of  $T_i$      $\pi_i$     Priority  
 $c_i$     Execution time of a job  
 $c'_i$     Time for entering critical section  
 $cs_i$     Execution time in critical section

## Traces of tasks

Time	process T <sub>1</sub>	process T <sub>2</sub>	process T <sub>3</sub>	process P
0	{}	{}	start?, {(cpu,1)}	P
1	{}	{}	req1, p1, {(cpu,1)} •	req?1, p?1, V(1)
2	{}	{}	{(cpu,1)}	• V(1)
3	{}	{}	{(cpu,1)}	• V(1)
4	{}	{}	{(cpu,1)}	• V(1)
5	start?, {(cpu,3)}	{}	{}	V(1)
6	{(cpu,3)}	{}	{}	V(1)
7	req3, {}	{}	chan?3, {(cpu,3)} •	req?3, chan3, V(3)
8	{}	{}	{(cpu,3)}	• V(3)
9	{}	{}	{(cpu,3)}	• V(3)
10	{}	start?, {}	{(cpu,3)}	• V(3)
11	{}	{}	{(cpu,3)}	• V(3)
12	{}	{}	{(cpu,3)}, v?	v!, P
13	p3, {(cpu,3)} •	{}	{}	p?3, V(3)
14	{(cpu,3)}, v?	{}	{}	v!, P
15	{(cpu,3)}	{}	{}	P
16	{(cpu,3)}	{}	{}	P
17	{}	{(cpu,2)}	{}	P
18	{}	{(cpu,2)}	{}	P
19	{}	{(cpu,2)}	{}	P
20	{}	{(cpu,2)}	{}	P
21	{}	{(cpu,2)}	{}	P
22	{}	req2, p2, {(cpu,2)} •	{}	req?2, p?2, V(2)
23	{}	{(cpu,2)}, v?	{}	v!P
24	{}	{(cpu,2)}	{}	P
25	{}	{}	{(cpu,1)}	P
26	{}	{}	{(cpu,1)}	P

(\*: in critical section)

5/27/08

Korea University

75

## Weak Bisimulation

Def. If  $t \in D^*$ , then  $\hat{t} \in (D - \{\tau\})^*$  is the sequence derived by deleting all occurrences of  $\tau$  from  $t$ .

Def. If  $t = \alpha_1 \dots \alpha_n \in D^*$ , then  $E \Rightarrow E'$  if  
 $P(\xrightarrow{(\tau, \_)}^*)^* \xrightarrow{\alpha_1} (\xrightarrow{(\tau, \_)}^*)^* \dots (\xrightarrow{(\tau, \_)}^*)^* \xrightarrow{\alpha_n} (\xrightarrow{(\tau, \_)}^*)^* P'$ ,  
 where "  $\_$  " in  $(\tau, \_)$  represents arbitrary integer.

Def. For a given transition system " $\rightarrow$ ", any binary relation  $r$  is a weak bisimulation if, for  $(P, Q) \in r$  and for any action  $\alpha \in D$ ,

1. if  $P \xrightarrow{\alpha} P'$ , then, for some  $Q', Q \xrightarrow{\alpha} Q'$  and  $(P', Q') \in r$ , and
2. if  $Q \xrightarrow{\alpha} Q'$ , then, for some  $P', P \xrightarrow{\alpha} P'$  and  $(P', Q') \in r$ .

Def.  $\approx_\tau$  is the largest weak bisimulation over " $\rightarrow_\tau$ ". It is an equivalence relation (though not a congruence) for ACSR.

5/27/08

Korea University

76

## Analyzing Real-Time Systems in ACSR-VP

- Two types of analyses
  - Validation
  - Schedulability analysis
- Basic idea
  - Checking weak bisimulation  $\approx_{\pi}$
  - Searching deadlocked states
- Practical Issues
  - Ensure that the **EDFSys** and **PIPSys** processes are finite state
  - Translate ACSR-VP processes to ACSR processes and use VERSA, the toolkit for ACSR

5/27/08

Korea University

77

## Validating the EDFSys Specification

Construct a correctness specification,  $EDFSpec$ , that is sequential and easy to inspect

Verify that  $EDFSys \approx_{\pi} EDFSpec$

$$\begin{array}{l}
 \overset{\text{def}}{EDFSpec} = [S(0, \dots, 0, 0)]_{\{cpu\}} \\
 \overset{\text{def}}{S}(s_1, t_1, \dots, s_n, t_n) = \\
 \left\{ \begin{array}{l}
 (s_i = c_i \wedge t_i = p_i) \\
 \quad \rightarrow (\tau, 1).S(\dots, s_{i-1}, t_{i-1}, 0, 0, s_{i+1}, t_{i+1}, \dots) \\
 + (s_i < c_i \wedge t_i = d_i) \\
 \quad \rightarrow (\tau, 1).NIL \\
 \sum_{1 \leq i \leq n} \left\{ \begin{array}{l}
 + (s_i = c_i \wedge t_i < p_i) \\
 \quad \rightarrow \emptyset : S(\dots, s_{i-1}, t_{i-1} + 1, s_i, t_i + 1, s_{i+1}, t_{i+1} + 1, \dots) \\
 + (s_i < c_i \wedge t_i < d_i) \\
 \quad \rightarrow \{(cpu, d_{max} - (d_i - t))\} \\
 \quad : S(\dots, s_{i-1}, t_{i-1} + 1, s_i + 1, t_i + 1, s_{i+1}, t_{i+1} + 1, \dots)
 \end{array} \right.
 \end{array} \right.
 \end{array}$$

5/27/08

Korea University

78

## Schedulability Analysis

---

Lemma 1 *If EDFSys is deadlock free, then it is schedulable.*

Lemma 2 *If*

$$\text{EDFSys} \setminus \{cpu\} \approx_{\pi} \emptyset^{\infty},$$

*then EDFSys is deadlock free.*

Lemma 3 *If PIPSys is deadlock free, then it is schedulable.*

Lemma 4 *If*

$$\text{PIPSys} \setminus \{cpu\} \approx_{\pi} \emptyset^{\infty},$$

*then PIPSys is deadlock free*

## Example 1

---

- Consider an instance  $\text{EDFSys}_1$  of  $\text{EDFSys}$  where:

Task  $T_1$ :  $c_1 = 1, d_1 = 2, p_1 = 3$

Task  $T_2$ :  $c_2 = 2, d_2 = 3, p_2 = 3$

- The following sufficient condition for schedulability from [Liu and Lay 73] is not satisfied:

$$\frac{c_1}{d_1} + \frac{c_2}{d_2} \leq 1$$

- The following equation  $\text{EDFSys} \setminus \{cpu\} \approx_{\pi} \emptyset^{\infty}$ ,

is satisfied, i.e., the task system is *schedulable*.

More specifically, we have

$$\begin{array}{ccccccc} \text{EDFSys1} & \xrightarrow{(\tau,2)}_{\pi} & \xrightarrow{(\tau,2)}_{\pi} & \xrightarrow{\{(cpu,2)\}}_{\pi} & \xrightarrow{\{(cpu,3)\}}_{\pi} & \xrightarrow{(\tau,3)}_{\pi} & \xrightarrow{\quad} \\ & & & & & & \\ & & & \xrightarrow{\{(cpu,3)\}}_{\pi} & \xrightarrow{(\tau,3)}_{\pi} & \text{EDFSys1} & \end{array}$$



