

Appunti sistemi embedded

Scheduling in sistemi real-time

2025

Contents

1	Introduzione	3
2	Real time	3
2.1	Limiti dei sistemi real-time correnti	4
2.2	Feature desiderabili di un sistema real-time	5
3	Fonti di imprevedibilità	5
3.1	DMA	6
3.2	Cache	6
3.3	Interrupts	6
3.3.1	Approccio A	6
3.3.2	Approccio B	7
3.3.3	Approccio C	7
3.4	Semafori	8
3.5	Gestione della memoria	8
3.6	Linguaggi di programmazione	8
4	Concetti di base di scheduling	8
4.1	Introduzione	8
4.2	Tipi di vincoli di scheduling	9
4.2.1	Vincoli temporali	10
4.2.2	Vincoli di precedenza	11
4.2.3	Vincoli di mutua esclusione	12
4.3	Classificazione di algoritmi di scheduling	13
4.3.1	Algoritmi di scheduling basati sulla garanzia	14
4.3.2	Algoritmi di scheduling best effort	15
5	Scheduling di task aperiodici	15
5.1	Garanzie	15
5.2	Earliest Deadline First (EDF)	16
5.2.1	Garanzia	16
5.3	EDF con vincoli di precedenza	16

6	Scheduling di task periodici	16
6.1	Fattore di utilizzo del processore	17
6.2	Timeline scheduling	18
6.3	Rate Monotonic Scheduling (RMS)	19

1 Introduzione

I sistemi real time sono sistemi di computazione dove la reazione a eventi dell'ambiente deve avvenire entro precisi requisiti temporali. Dunque la correttezza della computazione non dipende unicamente dal risultato prodotto, ma anche dal tempo in cui il risultato è prodotto. Una reazione ritardata potrebbe essere inutile o addirittura dannosa. Molto spesso questi sistemi vengono implementati attraverso la programmazione in assembly o con driver a basso livello per la gestione delle periferiche, manipolazione delle task e la priorità degli interrupt.

Questo approccio presenta degli svantaggi

- Programmazione tediosa
- Codice poco comprensibile
- Difficile manutenibilità
- Difficile verifica dei vincoli temporali

La conseguenza maggior di questo approccio è che il software costruito tramite tecniche empiriche può essere imprevedibile.

Se tutti i vincoli temporali critici non possono essere verificati a priori e il sistema operativo non include meccanismi specifici per maneggiare i task real-time il sistema potrebbe apparentemente funzionare correttamente per un certo periodo di tempo, ma potrebbe fallire in situazioni particolari. Il testing, pur essendo importante, non può essere una soluzione per ottenere la predicibilità in un sistema real-time. Questo vale perché il flusso di controllo dipende anche da eventi esterni dell'ambiente, che non possono essere interamente inclusi nel testing.

2 Real time

La correttezza del sistema non dipende solamente dal risultato logico prodotto dalla computazione ma anche dal tempo in cui il risultato è prodotto. Il termine reale indica che la reazione del sistema a eventi esterni deve avvenire **durante** la loro evoluzione. Di conseguenza il tempo interno al sistema deve essere misurato usando la stessa scala utilizzata per misurare il tempo dell'ambiente controllato. Al livello di processo la principale differenza tra un task real-time e non è la presenza di una **deadline**, ovvero il tempo massimo in cui una task deve essere completata. In un sistema critico una task che non rispetta la deadline non solo è in ritardo, è anche un errore! In base alle conseguenze a cui porta il non rispettare la deadline le task possono essere suddivise in:

- **Hard** se il non rispetto della deadline porta a conseguenze catastrofiche, ad esempio il fallimento di un sistema di controllo di un aereo.
- **Firm** se il non rispetto della deadline porta a un output inutile al sistema ma non produce conseguenze catastrofiche.
- **Soft** se il non rispetto della deadline porta a una degradazione della qualità del servizio ma non produce conseguenze catastrofiche.

Un sistema in grado di gestire task hard real-time è detto **hard real-time system**. Alcuni di questi sistemi sono tendenzialmente sistemi **safety critical**.

2.1 Limiti dei sistemi real-time correnti

Molti dei sistemi attualmente in uso sono soluzioni basate su kernel, che sono versioni modificate di sistemi operativi a time sharing. Una conseguenza di questo è che condividono le stesse feature di base dei sistemi a time sharing, che non sono adatte a sistemi real-time. Le caratteristiche principali sono:

- **Multitasking**

Supporto per la programmazione concorrente offerta attraverso un insieme di chiamate a sistema per la gestione dei processi. Molte di queste primitive non prendono in considerazione il tempo e, peggio ancora, introducono ritardi senza limiti ai tempi di esecuzione, il che potrebbe portare le task hard a non rispettare le deadline in maniera non predicibile.

- **Scheduling basato su priorità**

Meccanismo flessibile di scheduling, dato che permette di implementare diverse strategie in base alla maniera in cui assegno le priori. Nonostante ciò, mappare i constraint a un insieme di priorità è difficile e non sempre possibile, specialmente in un sistema dinamico. Il problema principale è che i kernel permettono un **numero limitato di livelli di priorità**, mentre le deadline possono variare molto di più. In sistemi dinamici l'arrivo di una nuova task può comportare un remappaggio dell'intero insieme di priorità.

- **Capacità di rispondere velocemente agli interrupt**

Solitamente questa feature viene ottenuta tramite la possibilità di impostare una priorità per gli interrupt, più alta rispetto alla priorità dei processi e riducendo al minimo le porzioni di codice eseguite con le interruzioni disabilitate.

Nota che questo approccio aumenta la reattività del sistema agli eventi esterni, ma introduce dei ritardi non limitati all'esecuzione dei processi. Infatti, un processo di un'applicazione sarà sempre interrotto da un interrupt di un driver, anche se è più importante che venga servita la periferica.

- **Meccanismo di base per la comunicazione tra processi e sincronizzazione**

Tipicamente semafori binari per la sincronizzazione e la mutua esclusione su risorse condivise. Però, se nessun protocollo d'accesso è utilizzato per entrare in sezioni critiche i semafori possono generare una serie di problemi, ad esempio la **priority inversion** o deadlock.

- **Small kernel e context switch veloce**

Questa feature riduce l'overhead del sistema, dunque migliora il response time medio dell'insieme di task. Ciò nonostante non è sufficiente per garantire il rispetto delle deadline. D'altro canto un kernel piccolo implica limitate funzionalità, il che ha effetto sulla predicibilità del sistema.

- **Supporto a clock real-time come reference di tempo interna**

Feature essenziale di ogni kernel real-time che gestisce task hard real-time che interagiscono con l'ambiente esterno. Nella maggior parte dei kernel commerciali questo è l'unico meccanismo di temporizzazione disponibile. In molti casi non vi sono primitive per controllare esplicitamente i constraint temporali delle task, e non c'è un supporto per l'attivazione automatica delle task periodiche.

La mancanza di una qualsiasi garanzia preclude l'utilizzo di questi sistemi in casi in cui i requisiti temporali siano stringenti.

2.2 Feature desiderabili di un sistema real-time

La predicibilità può essere ottenuta apportando pesanti modifiche ai tipici sistemi a time sharing. Alcune proprietà importanti sono:

- **Timeliness**

I risultati non devono essere corretti solo nel loro valore ma anche nel dominio temporale. Di conseguenza il sistema operativo deve fornire degli specifici meccanismi del kernel per la gestione del tempo e per la gestione dei task con constraint temporali espliciti e criticità diverse.

- **Predicibilità**

Per raggiungere il livello di performance desiderato, il sistema deve essere analizzabile per prevedere le conseguenze di ogni decisione di scheduling. Nelle applicazioni safety critical tutti i requisiti temporali devono essere garantiti a priori, prima di mettere il sistema in azione. Se qualche vincolo non può essere garantito, il sistema deve essere in grado di informare l'utente e delle decisioni alternative devono essere prese.

- **Efficienza**

La maggior parte dei sistemi real-time sono embedded, dunque l'efficienza è un requisito fondamentale.

- **Robustezza**

Il sistema deve essere in grado di continuare a funzionare a qualsiasi condizioni di carico. La gestione dell'overload e il comportamento di adattamento sono feature essenziali per sistemi con molta varianza di carico.

- **Tolleranza agli errori**

Singoli errori hardware o software non devono portare a un fallimento del sistema.

- **Manutenibilità**

L'architettura di un sistema real-time deve essere progettata in maniera modulare per permettere di eseguire modifiche facilmente.

3 Fonti di imprevedibilità

Un sistema dovrebbe essere in grado di prevedere l'evoluzione delle task e garantire in anticipo che tutti i vincoli temporali critici vengano rispettati. L'affidabilità della garanzia dipende da un insieme di fattori, che includono anche le caratteristiche fisiche dell'hardware e i meccanismi/politiche del kernel.

La prima componente che influisce sulla predicibilità è il processore. Le caratteristiche del processore, ad esempio il prefetch delle istruzioni, pipelining, cache, DMA sono fonti di **non determinismo**.

Pur aumentando le performance medie del sistema, introducono fattori di non determinismo che prevengono una precisa stima dei worst-case-execution-time (WCETs). Altre caratteristiche che influenzano la predicibilità sono le caratteristiche interne del kernel,

come l'algoritmo di scheduling, i meccanismi di sincronizzazione, i tipi di semafori, le politiche di gestione della memoria e le politiche di gestione degli interrupt.

3.1 DMA

Uno dei modi più comuni per implementarlo è tramite cycle stealing, il DMA controller prende il controllo del bus per un ciclo di clock e poi lo rilascia al processore. Durante l'operazione di DMA il trasferimento e l'esecuzione del codice sul processore avvengono in parallelo. Se la CPU e il DMA controller richiedono un ciclo di memoria allo stesso tempo, il bus viene assegnato al DMA controller e la CPU aspetta finché il DMA non ha finito. In questo modo non è possibile prevedere il tempo di esecuzione di un task, dato che il tempo di esecuzione dipende da quante volte il DMA controller ha bisogno del bus. Una possibile soluzione al problema è il time slicing, dove un ciclo di memoria è diviso in due slot di tempo adiacenti, uno per il DMA e uno per la CPU. Questa soluzione è più costosa ma più prevedibile di cycle stealing, dato che il response time di una task non è influenzato dal DMA.

3.2 Cache

Nei sistemi real-time la cache è una fonte di non determinismo, dato che il tempo di accesso alla memoria dipende dalla presenza o meno del dato in cache. Quando avviene un cache miss il tempo di accesso alla memoria è molto più lungo, e questo non è prevedibile. Quando si effettua una scrittura in memoria l'utilizzo della cache risulta costoso perché il tempo di accesso alla memoria è più lungo, dato che deve essere garantita la consistenza.

3.3 Interrupts

Un grosso problema per la predicibilità è la gestione degli interrupt provenienti dalle periferiche I/O. Se non gestite correttamente possono introdurre ritardi non limitati durante l'esecuzione di processi.

In quasi tutti i sistemi operativi la gestione di un interrupt prevede l'esecuzione di una service routine dedicata per la gestione del device. Il vantaggio di questo metodo è che tutti i dettagli hardware sono contenuti nel driver.

In molti OS gli interrupt vengono serviti usando un sistema a priorità fissa, tramite i quali ogni driver viene schedato in base a una priorità statica, questa priorità è più alta rispetto a quella di qualsiasi processo. In generale è molto difficile prevedere a priori il numero di interrupt che un task può ricevere, dunque il delay è imprevedibile.

Per ridurre l'interferenza dei driver sui task ed effettuare operazioni di I/O le periferiche devono essere gestite diversamente

3.3.1 Approccio A

La soluzione più radicale è disattivare tutti gli interrupt esterni, a eccezione di quelli generati dal timer. Tutte le periferiche vanno gestite dall'applicazione (controllo di programma), i quali hanno accesso ai registri delle periferiche. Dato che non ci sono più interrupt, il trasferimento deve essere effettuato tramite polling. L'accesso diretto alle periferiche consente alta flessibilità al livello di programma ed elimina i delay introdotti dai driver. Dunque il tempo necessario per un trasferimento può essere calcolato a priori. Un altro vantaggio è che il kernel non va modificato quando una nuova periferica viene

aggiunta o ne viene rimossa una.

Lo svantaggio principale è che risulta poco efficiente dato la busy wait necessaria per il polling. Un grosso problema è che le applicazioni richiedono la conoscenza completa dei dettagli a basso livello delle periferiche che devono gestire. Questo può essere gestito incapsulando le routine che dipendono dai device in una libreria.

3.3.2 Approccio B

Come nella soluzione precedente tutti gli interrupt esterni sono disabilitati, tranne quelli dei timer.

In questa soluzione le periferiche non vengono gestite direttamente dall'applicazione, ma a turno da routine dedicate del kernel periodicamente attivate dal timer. Questo approccio elimina i ritardi causati dai driver e confina le operazioni di I/O a una o più routine periodiche del kernel. In alcuni sistemi i device I/O vengono suddivisi in due gruppi in base alla loro velocità, quelli più veloci vengono gestiti da task periodici con frequenza maggiore.

Il vantaggio di questo metodo è che tutti i dettagli hardware sono incapsulati nel kernel, dunque l'applicazione non ha bisogno di conoscere i dettagli a basso livello delle periferiche. Dato che gli interrupt sono disabilitati il maggior problema è dato dalla busy wait delle procedure di gestione dell'I/O del kernel. Questo metodo introduce un overhead maggiore rispetto all'approccio A, richiede modifiche al kernel quando una periferica viene aggiunta o rimossa.

3.3.3 Approccio C

Lasciare gli interrupt abilitati e ridurre al minimo possibile i driver per le periferiche. L'unico scopo dei driver è quello di attivare una task che gestisce il device. Una volta attivato il task di gestione del device viene eseguito sotto diretto controllo del sistema operativo ed è schedulato come un task normale. In questo modo la priorità che può essere assegnata al task di gestione del device è completamente indipendente dalle altre priorità può essere impostata in base ai requisiti del programma.

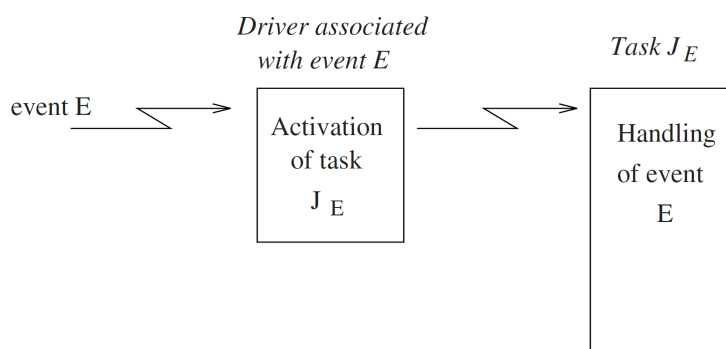


Figure 1: Attivazione task per la gestione del device

Il vantaggio principale di questo metodo è che il busy wait viene eliminato e il tempo di esecuzione del driver di gestione del device è molto ridotto, di conseguenza il tempo di esecuzione del task è più prevedibile. Nella maggior parte delle applicazioni pratiche il piccolo overhead portato da questo tipo di driver viene ignorato.

3.4 Semafori

L'usuale meccanismo dei semafori utilizzato nei sistemi operativi tradizionali non è adatto per i sistemi real-time per via del fenomeno dell'**inversione di priorità**. L'inversione di priorità deve essere evitato a tutti i costi in un sistema real-time, dato che introduce non determinismo sull'esecuzione di task critiche. Per quanto riguarda il problema della mutua esclusione l'inversione di priorità può essere evitata usando protocolli specifici che devono essere usati ogni volta che si effettua un accesso alla sezione critica. L'implementazione di questo tipo di protocollo potrebbe richiedere estensive modifiche al kernel, che non comprendono solo le primitive *wait* e *signal*, ma anche strutture dati e meccanismi di task management.

3.5 Gestione della memoria

Come gli altri aspetti del kernel, la gestione della memoria deve essere progettata per evitare il non determinismo. Ad esempio sistemi di paging on-demand non sono adatti a sistemi con rigidi vincoli temporali a causa dei lunghi e imprevedibili ritardi causati dalle page fault e page replacement. Le soluzioni tipiche adottate nei sistemi real-time prevedono regole di segmentazione della memoria con uno schema fisso. Il partizionamento statico è particolarmente efficace quando i programmi richiedono una quantità di memoria simile tra loro. In generale gli schemi di allocazione statici aumentano la predicibilità del sistema, ma diminuiscono la flessibilità in sistemi dinamici, dunque la scelta progettuale va fatta bilanciando le due caratteristiche.

3.6 Linguaggi di programmazione

Con l'aumentare della complessità dei sistemi real-time aumenta anche la necessità di utilizzare astrazioni offerte dai linguaggi.

Purtroppo i linguaggi attualmente a disposizione non sono abbastanza espressivi da permettere di esprimere certi comportamenti temporali, dunque non sono adatti per la programmazione di sistemi real-time. L'esistenza di costrutti che modificano il flusso di controllo del programma in maniera non deterministica, come gli *if*, rende impossibile la stima affidabile dei WCETs in attività concorrenti. Un'altra fonte di imprevedibilità è data dalla ricorsione, l'analizzatore di scheduling non è in grado di prevedere il numero di chiamate ricorsive e il tempo di esecuzione.

4 Concetti di base di scheduling

4.1 Introduzione

Un **processo** è una computazione eseguita dalla CPU in maniera sequenziale, d'ora in poi useremo il termine processo come sinonimo di *task* o *thread*.

In realtà un processo è una entità più complessa composta da vari task (o thread) che condividono lo stesso spazio di indirizzamento e le stesse risorse. Una **politica di scheduling** è un criterio secondo il quale la CPU viene assegnata ai vari task che si sovrappongono. L'insieme di politiche che in ogni istante di tempo determinano l'ordine con il quale i task vengono eseguiti è detto **algoritmo di scheduling**.

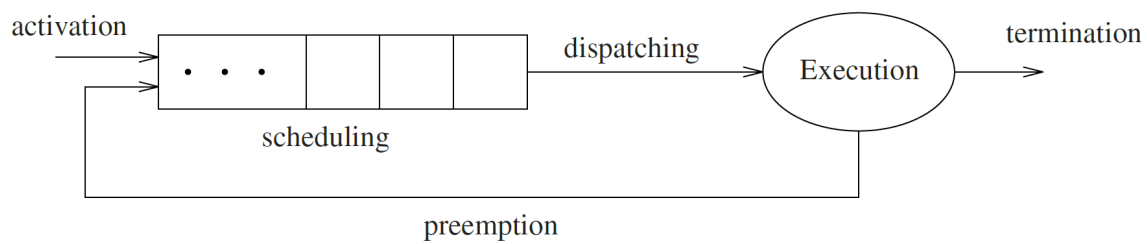


Figure 2: Coda di task in ready che aspettano di essere eseguiti

L'operazione specifica di assegnare la CPU a un task selezionato dall'algoritmo di scheduling è detta **dispatching**. Quando un processo può essere potenzialmente eseguito dalla CPU, possiamo avere due casi:

- Il processo è in **running** perché è già stato selezionato dall'algoritmo di scheduling e la CPU è stata assegnata a lui.
- Il processo è in **ready** e aspetta di essere selezionato dall'algoritmo di scheduling.

In entrambi i casi il processo viene detto un task **attivo**. Tutti i processi in wait vengono tenuti in una **ready queue**. I vari sistemi operativi possono utilizzare anche più code dipendentemente dal tipo di task. In molti sistemi dove è concessa l'attivazione dinamica dei task, questi possono essere interrotti in qualsiasi momento, così che quando un task più importante entra nel sistema esso possa essere eseguito immediatamente senza aspettare nella ready queue. L'operazione di interrompere il task in esecuzione e reinserirlo nella ready queue è detta **preemption**.

Nei sistemi real-time la preemption è importante per 3 motivi:

- I task che gestiscono le eccezioni potrebbero aver bisogno di fare preemption di task in esecuzione per garantire che le eccezioni siano gestite in tempo.
- Quando i task hanno livelli di priorità diversi, i task critici possono essere eseguiti immediatamente.
- Tipicamente lo scheduling con preemption permette maggiore efficienza, nel senso che permette l'esecuzione di task real time con una percentuale di utilizzo di CPU maggiore.

D'altro canto la preemption distrugge la località dei programmi e introduce overhead, di conseguenza limitare la preemption in un sistema real-time può essere benefico.

Uno schedule è detto **feasible** se tutti i task possono essere eseguiti rispettando l'insieme di vincoli a essi posti. Un insieme di task è detto **schedulabile** se esiste uno schedule che rispetta i vincoli di tutti i task.

4.2 Tipi di vincoli di scheduling

Tipicamente i vincoli applicabili a task real time rientrano in tre classi:

- Vincoli temporali

- Vincoli di precedenza
- Vincoli di mutua esclusione su risorse condivise

4.2.1 Vincoli temporali

I sistemi realtime sono caratterizzati da attività computazionali con stringenti vincoli temporali che devono essere rispettati per ottenere il comportamento desiderato. Un requisito tipico è che un task deve essere completato entro un certo tempo, detto **deadline**. Se la deadline è rappresentata in rispetto a quando il task è stato attivato, essa è detta **relative deadline**, altrimenti è detta **absolute deadline**.

In base alle conseguenze a cui porta il non rispetto della deadline le task possono essere suddivise in:

- **Hard** se il non rispetto della deadline porta a conseguenze catastrofiche, ad esempio il fallimento di un sistema di controllo di un aereo.
- **Firm** se il non rispetto della deadline porta a un output inutile al sistema ma non produce conseguenze catastrofiche.
- **Soft** se il non rispetto della deadline porta a una degradazione della qualità del servizio ma non produce conseguenze catastrofiche.

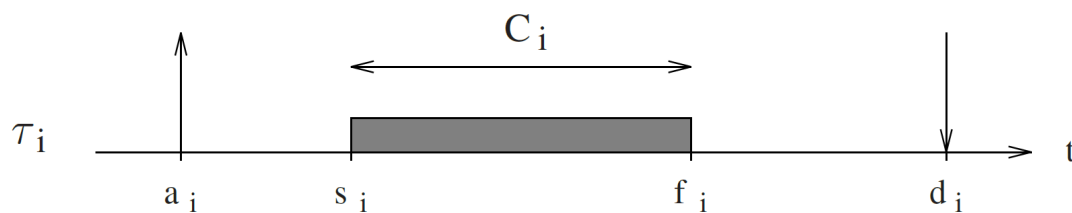


Figure 3: Parametri di un task

In generale un task realtime τ_i è caratterizzato da:

- **Tempo di arrivo** a_i , il momento in cui il task è pronto per essere eseguito, anche detto **release time** o **request time** e indicato da r_i .
- **Tempo di computazione** c_i , il tempo necessario per completare il task senza interruzioni, l'upperbound è il WCET.
- **Deadline assoluta** d_i , il tempo entro il quale il task deve essere completato.
- **Deadline relativa** D_i , il tempo entro il quale il task deve essere completato rispetto al suo tempo di arrivo, $D_i = d_i - r_i$.
- **Tempo di inizio** s_i , il tempo in cui il task inizia a essere eseguito.
- **Tempo di completamento** f_i , il tempo in cui il task termina la sua esecuzione.
- **Tempo di risposta** T_i , il tempo di ritardo del task, ovvero $T_i = f_i - r_i$.

- **Criticità**, già discusso: hard, firm o soft.
- **Lateness** L_i , il tempo di ritardo del task rispetto alla deadline, ovvero $L_i = f_i - d_i$, in generale vogliamo che sia negativa, ovvero i task devono essere eseguiti entro le deadline.
- **Tempo di slack** $X_i = d_i - a_i - c_i$, il tempo massimo per cui l'attivazione di un task può essere ritardata senza violare la deadline.

Un'altra caratteristica temporale che riguarda la regolarità dell'attivazione di un task. Un task è detto **periodico** se consiste di una infinita sequenza di attività identiche, chiamate **job**, che si attivano a intervalli regolari. Per chiarezza di notazione notiamo i task periodici con τ_i e quelli aperiodici con J_i . Il generico k-esimo job di un task periodico τ_i è denotato con $\tau_{i,k}$. Il tempo di attivazione della prima istanza di un task periodico ($\tau_{i,1}$) è detto **fase**, denotato con ϕ_i . Il periodo di attivazione di un task periodico è denotato con T_i , che è il tempo tra due attivazioni consecutive. I task **aperiodici** sono caratterizzati da un tempo di attivazione non regolare, un task **sporadico** è un task aperiodico che ha un tempo di attivazione massimo, ovvero il tempo minimo tra due attivazioni consecutive.

4.2.2 Vincoli di precedenza

In alcune applicazioni le attività non possono essere eseguite in ordine arbitrario, ma devono rispettare un ordine di precedenza stabilito durante la progettazione. Tali precedenze tendenzialmente vengono rappresentate tramite un grafo diretto aciclico (DAG), dove i nodi rappresentano le attività e gli archi rappresentano le precedenze. Un DAG definisce un relazione di ordine parziale tra le attività, che può essere usata per determinare un ordine di esecuzione.

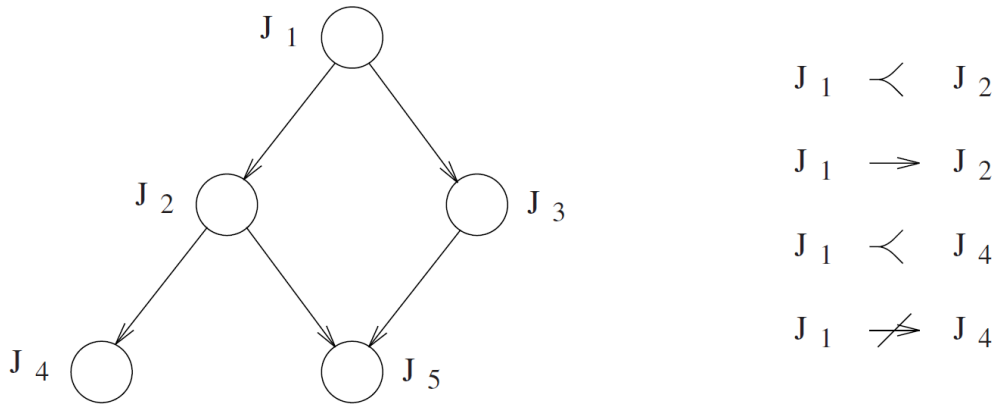


Figure 4: Relazione di precedenza tra 5 task

La notazione $J_a \prec J_b$ indica che il task J_a è un predecessore di J_b , ovvero il grafo contiene un cammino orientato da J_a a J_b .

La notazione $J_a \rightarrow J_b$ indica che J_a è un predecessore diretto di J_b , ovvero il grafo contiene un arco diretto da J_a a J_b .

I task senza predecessori vengono detti **task iniziali**, mentre quelli senza successori vengono detti **task finali**.

4.2.3 Vincoli di mutua esclusione

Dal punto di vista di un processo, una risorsa è una struttura software che può essere utilizzata dal processo per avanzare la sua esecuzione. Una risorsa è detta *privata* se è dedicata a un processo, altrimenti se può essere utilizzata da più processi è detta **condivisa**. Per mantenere la correttezza del dato, molte risorse condivise non permettono l'accesso simultaneo da parte di task concorrenti, dunque richiedono la **mutua esclusione**. La parte di codice che accede alla risorsa condivisa in mutua esclusione è detta **sezione critica**.

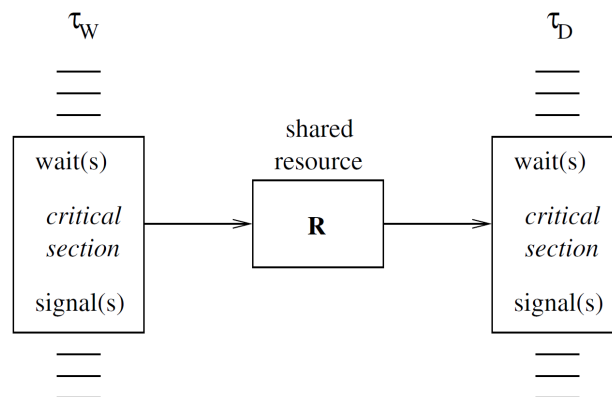


Figure 5: Struttura di due task che accedono a una risorsa condivisa in mutua esclusione, tramite meccanismo di semafori

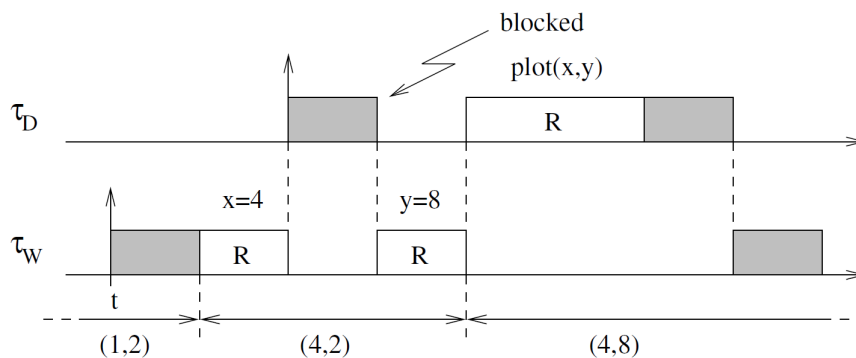


Figure 6: Esempio di scheduling quando la risorsa è protetta da un semaforo

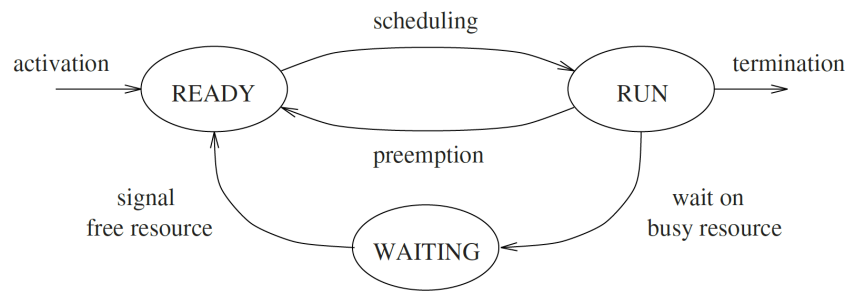


Figure 7: Stato di wait causato da un vincolo sulle risorse

Quando un task è in attesa di una risorsa condivisa, si dice che è *bloccato* su quella risorsa. Tutti i task bloccati su una stessa risorsa vengono messi in una **wait queue** associata a un semaforo che protegge la risorsa. Quando un task in esecuzione utilizza la primitiva *wait* su un semaforo bloccato, entra nello stato di **wait**, fino a quando un altro task non esegue la primitiva *signal* che sblocca il semaforo. Nota che quando un task lascia lo stato di wait non entra immediatamente nello stato di running, ma entra nello stato di ready, cosicché la CPU possa essere assegnata a un task con priorità più alta dall'algoritmo di scheduling.

4.3 Classificazione di algoritmi di scheduling

Le classi principali di algoritmi di scheduling sono:

- **Preemptive vs. non preemptive**

- Preemptive: un task può essere interrotto in qualsiasi momento da un task con priorità più alta, secondo una policy di scheduling definita in precedenza.
- Non preemptive: un task una volta avviato deve essere eseguito fino al termine della sua esecuzione. Tutte le scelte di scheduling vengono effettuate quando il task termina la sua esecuzione.

- **Static vs. Dynamic**

- Static: le decisioni di scheduling avvengono in base a un insieme di parametri che vengono definiti prima dell'attivazione dei task.
- Dynamic: le decisioni di scheduling avvengono in base a un insieme di parametri che possono essere modificati durante l'evoluzione del sistema.

- **Off-line vs. online**

- Off-line: Le decisioni di scheduling vengono interamente prese per l'intero insieme di task prima dell'attivazione di essi. Questo scheduling è memorizzato in una tabella e successivamente viene eseguito dal dispatcher.
- Online: Le decisioni di scheduling vengono prese a runtime ogni volta che un nuovo task entra nel sistema, o quando un task running termina. Notare che questo genera un overhead.

• Ottimale vs. euristico

- Ottimale: Un algoritmo è detto ottimale se minimizza la "funzione costo" definita per l'insieme di task. Quando non è definita questa funzione costo il nostro unico obiettivo è ottenere uno schedule feasible (se esiste).
- Euristico: Un algoritmo è detto euristico se è guidato da una funzione euristica per prendere decisioni di scheduling. Un algoritmo euristico tende verso lo scheduling ottimale, ma non è garantito che lo raggiunga.

Inoltre un algoritmo è detto *chiaroveggente* se conosce il futuro, ovvero conosce in anticipo il tempo di arrivo di tutti i task. Questo algoritmo chiaramente non esiste, ma è utilizzato come benchmark per valutare gli altri algoritmi in confronto al caso perfetto.

4.3.1 Algoritmi di scheduling basati sulla garanzia

Nelle applicazioni hard real-time sono richiesti comportamenti altamente predicibili, la feasibility di uno scheduling dovrebbe essere garantita in anticipo (prima dell'attivazione dei task). In questa maniera, e un task critico non può essere schedulato prima della sua deadline, il sistema è in tempo per eseguire un'azione alternativa che tenta di evitare conseguenze catastrofiche. Per controllare la feasibility di uno scheduling prima dell'esecuzione dei task il sistema deve pianificare le sue azioni assumendo il caso peggiore. Quindi vanno analizzati i parametri del task set per verificare che esista uno scheduling ammissibile. Nei sistemi statici, ovvero il task set è fisso e noto a priori, tutte le attivazioni dei task possono essere pre calcolate off-line, e tutto lo schedule può essere memorizzato in una tabella che contiene tutti i task garantiti. Poi a runtime il dispatcher può semplicemente eseguire lo schedule memorizzato. Il vantaggio principale dell'approccio statico è che l'overhead a runtime non dipende dalla complessità dell'algoritmo di scheduling. Questo permette di utilizzare algoritmi di scheduling sofisticati per risolvere problemi complessi. D'altro canto, il sistema è rigido rispetto a cambiamenti nell'ambiente.

Nei sistemi dinamici (che tipicamente consistono di task *firm*) i task possono essere creati a runtime, dunque la garanzia di scheduling deve essere effettuata online ogni volta che un nuovo task viene creato.

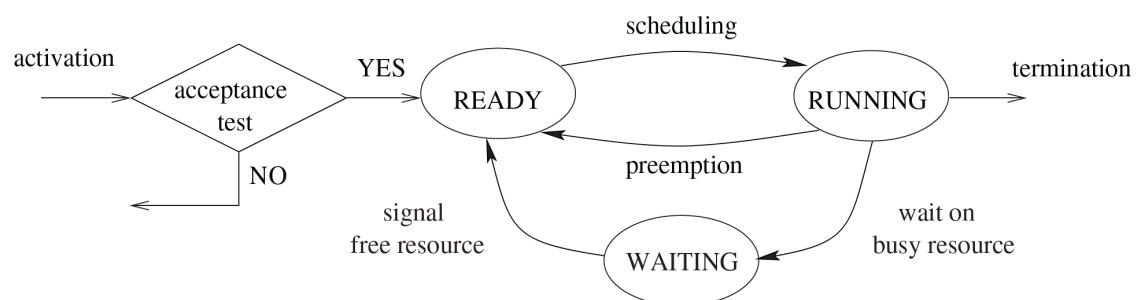


Figure 8: Meccanismo di garanzia in un sistema dinamico real time

Inoltre deve essere garantito che quando un task viene accettato, esso non deve interdire la feasibility di task garantiti in precedenza.

4.3.2 Algoritmi di scheduling best effort

Nelle applicazioni real-time soft, i task non hanno vincoli temporali stringenti, ma devono essere eseguiti in un tempo ragionevole per soddisfare i requisiti si sistema. Dunque il mancato rispetto delle deadline non porta a conseguenze catastrofiche, ma porta a una degradazione delle performance. Un algoritmo best effort cerca di fare "del suo meglio" per rispettare le deadline, ma non garantisce di trovare uno scheduling feasible. Negli approcci best effort i task potrebbero essere messi in coda in base a delle politiche che considerano i requisiti temporali, ma non viene controllata la feasibility e quindi un task potrebbe venir abortito durante la sua esecuzione.

Mediamente le performance di un algoritmo best effort sono migliori rispetto a quelle di un algoritmo basato sulla garanzia, infatti un algoritmo basato sulla garanzia potrebbe, basandosi sul caso peggiore, rifiutare inutilmente un task che in realtà sarebbe riuscito a rispettare la deadline.

5 Scheduling di task aperiodici

Gli algoritmi affrontati in questa sezione riguardano task aperiodici real-time, eseguiti su una macchina single core. Ogni algoritmo è una soluzione a un problema di scheduling espresso come insieme di assunzioni sulle task e criteri di ottimalità da utilizzare. Pur essendo pensati per task aperiodici su macchine monoprocesso, molti di questi algoritmi possono essere estesi per lavorare su sistemi multiprocessore o architetture distribuite per lavorare con task set più complessi.

Per descrivere un problema di scheduling utilizziamo i parametri $\alpha|\beta|\gamma$:

- α è il tipo di macchina (uniprocessore, multiprocessore, distribuito, ecc.)
- β descrive le task e le caratteristiche delle risorse (preemptive, indipendente vs vincoli di precedenza, attivazione asincrona, etc.)
- γ è il criterio di ottimalità

5.1 Garanzie

Per garantire che esista uno scheduling feasible dobbiamo mostrare che, nel caso peggiore, tutti i task vengano completati prima delle rispettive deadline. Vale a dire che per ogni task dobbiamo mostrare che il tempo di completamento nel caso peggiore f_i è minore o uguale alla rispettiva deadline d_i .

$$f_i \leq d_i, \forall i \in \{1, 2, \dots, n\} \quad (1)$$

Se un task ha requisiti hard, quest'analisi di fattibilità deve essere effettuata prima dell'esecuzione dei task. Senza perdita di generalità possiamo assumere che i task J_1, J_2, \dots, J_n siano elencati in ordine di deadline crescente, ovvero J_1 ha la deadline più vicina. In questo caso il tempo di completamento del task J_i nel caso peggiore è dato da:

$$f_i = \sum_{k=1}^i C_k \quad (2)$$

Dunque in un task set con n task il test di garanzia consiste nel verificare che:

$$\forall i \in \{1, 2, \dots, n\} \sum_{k=1}^i C_j \leq d_i \quad (3)$$

5.2 Earliest Deadline First (EDF)

Se le task non sono sincrone ma possono avere tempo di arrivo arbitrario, la preemption diventa un fattore importante. In generale un problema di scheduling dove la preemption è permessa è sempre più facile da risolvere rispetto allo stesso problema senza preemption. Una soluzione elegante a problemi del tipo $1|preem|L_{max}$ è data dall'algoritmo **Earliest Deadline First** (EDF).

Theorem 1 *Dato un insieme di n task con tempo di arrivo arbitrario, ogni algoritmo che ad ogni istante esegue il task con la deadline assoluta più vicina tra tutti i task attivi è ottimale rispetto alla minimizzazione della lateness.*

La complessità computazionale dell'algoritmo è $O(n)$ per ogni task se la ready queue è implementata come lista, e $O(n \log n)$ se la ready queue è implementata come heap.

5.2.1 Garanzia

Quando i task hanno attivazioni dinamiche e il tempo di arrivo non è noto a priori, il test di garanzia deve essere fatto dinamicamente ogni volta che un task entra nel sistema. Sia \mathcal{T} l'insieme corrente di task attivi, che sono stati garantiti in precedenza, sia J_{new} il nuovo task che entra nel sistema. Per poter accettare J_{new} dobbiamo garantire che il nuovo task set $\mathcal{T} \cup J_{new}$ sia schedulabile.

5.3 EDF con vincoli di precedenza

Consideriamo problemi di scheduling del tipo $1|prec, preem|L_{max}$, dove i task sono soggetti a vincoli di precedenza. Il problema di schedulare un insieme di n task con vincoli di precedenza con attivazioni dinamiche può essere risolto in tempo polinomiale SOLO se i task sono preentable. L'idea alla base dell'approccio è quella di trasformare il set di task dipendenti \mathcal{T} in un insieme di task indipendenti \mathcal{T}^* tramite adeguate modifiche ai parametri temporali, dopodiché possiamo applicare l'algoritmo EDF. L'algoritmo di trasformazione garantisce che \mathcal{T} sia schedulabile se e solo se \mathcal{T}^* è schedulabile. In pratica tutti i release time e le deadline sono modificati in modo che ogni task non possa cominciare prima del predecessore e non possa fare preemption su i suoi successori.

6 Scheduling di task periodici

In molti sistemi real-time di controllo, le attività periodiche compongono la maggior parte della computazione del sistema. Quando una applicazione di controllo consiste in molteplici task periodici concorrenti, con vincoli temporali indipendenti, il sistema operativo deve garantire che ogni istanza periodica venga regolarmente attivata con la giusta frequenza e che venga eseguita entro la deadline (che in generale è diversa dal suo periodo).

Per semplificare l'analisi di scheduling vengono assunte le seguenti ipotesi:

- Le istanze di task periodici τ_i sono regolarmente attivate a frequenza costante, ovvero ogni T_i (periodo del task).
- Tutte le istanze di un task periodico τ_i hanno lo stesso WCET C_i .
- Tutte le istanze di un task periodico τ_i hanno la stessa deadline relativa D_i , che è uguale a T_i .
- Tutte le task di \mathcal{T} sono indipendenti.

Inoltre alcune assunzioni implicite sono:

- Nessun task si può bloccare da solo, ad esempio per operazioni di I/O.
- Tutti i task vengono rilasciati appena arrivano nel sistema.
- L'overhead del kernel viene considerato nullo.

Per quest'ultima assunzione potrei evitare problemi se considero solo una percentuale di "potenza di cpu" disponibile, così da lasciare dello spazio per l'overhead del kernel.

Nota che le prime due assunzioni non sono restrittive, dato che in molte applicazioni di controllo ogni attività periodica richiede l'esecuzione della stessa routine a intervalli regolari. Quando si analizzano casi realistici, le assunzioni 3 e 4 vengono rilassate.

Altri parametri importanti per i task periodici sono:

- **Hyperperiod**: minimo intervallo di tempo nel quale lo scheduling si ripete.
- **Critical instant** di un task: tempo di arrivo che produce il response time del task più lungo.

6.1 Fattore di utilizzo del processore

Dato un insieme di task periodici Γ , il *fattore di utilizzo* del processore U è la frazione del tempo di CPU che viene utilizzato per eseguire l'insieme di task. Dato che $\frac{C_i}{T_i}$ è la frazione di tempo di CPU utilizzata da un task τ_i , il fattore di utilizzo del processore è dato da:

$$U = \sum_{i=1}^n \frac{C_i}{T_i} \quad (4)$$

Questo valore ci fornisce una misura del carico computazionale sulla CPU. Nonostante possa essere migliorato aumentando il tempo di computazione dei task o diminuendo il loro periodo, esiste un valore massimo di U per il quale Γ è schedulabile. Tale limite dipende dalla relazione dei periodi dei task e dall'algoritmo di scheduling utilizzato. Sia $U_{ub}(\Gamma, A)$ il limite superiore al fattore di utilizzo del processore per un insieme di task Γ schedulato con l'algoritmo A . Quando $U = U_{ub}(\Gamma, A)$, si dice che Γ *utilizza completamente* il processore. In questo caso Γ è schedulabile da A , ma un qualsiasi aumento nel tempo di computazione rende il task set non feasible. Per un algoritmo di scheduling A , il limite superiore minore $U_{lub}(\Gamma, A)$ è il minimo fattore di utilizzo su tutti i task set che utilizzano completamente il processore.

$$U_{lub}(\Gamma, A) = \min_{\Gamma} U_{ub}(\Gamma, A) \quad (5)$$

Dato che $U_{lub}(\Gamma, A)$ è il minimo di tutti gli upper bound, qualsiasi task set con fattore di utilizzo minore di $U_{lub}(\Gamma, A)$ è sicuramente schedulabile da A . In questa maniera è facilmente determinabile se un task set è schedulabile o meno con un determinato algoritmo.

- Se $U_{lub} < U < 1.0$, la schedulabilità viene raggiunta solo se i periodi sono compatibili.
- Se $U_{lub} > 1.0$, il task set non è schedulabile da nessun algoritmo.

6.2 Timeline scheduling

Anche noto come *Cyclic Executive*, è uno degli approcci più adottati per gestire task periodici nei sistemi di difesa militari e di controllo di traffico. Il metodo consiste nel dividere l'asse temporale in slot di egual lunghezza, nei quali uno o più task vengono allocati per l'esecuzione; in modo da rispettare le frequenze date dai requisiti. Un timer sincronizza l'attivazione dei task all'inizio di ogni slot.

Consideriamo un esempio dove tre task A, B e C devono essere eseguiti con frequenze rispettivamente di $40Hz$, $20Hz$ e $10Hz$. Dunque i rispettivi periodi sono $T_A = 25ms$, $T_B = 50ms$ e $T_C = 100ms$ (N.B. $F = \frac{1}{T}$), si ricava che la lunghezza ottimale dello slot è di $25ms$, che è il massimo comune divisore dei periodi. Quindi il task A viene eseguito ogni slot, il task B ogni 2 slot e il task C ogni 4 slot.

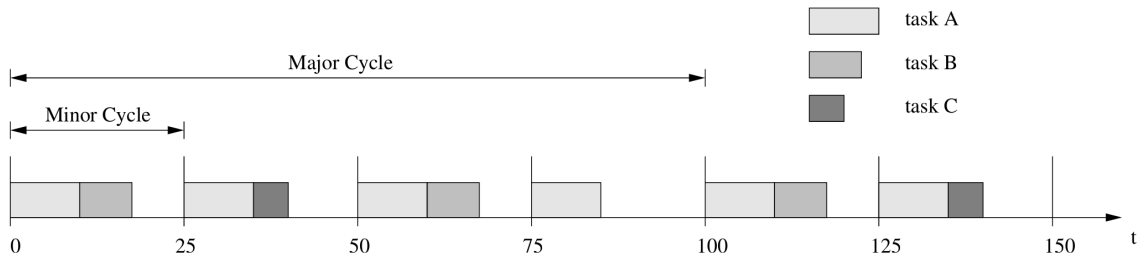


Figure 9: Esempio di timeline scheduling

La durata dello slot è detta *Minor cycle*, mentre l'intervallo minimo di tempo dopo il quale uno schedule si ripete è detto *Major cycle*.

In generale, il major cycle è uguale al minimo comune multiplo di tutti i periodi. Per poter garantire a priori che lo schedule sia feasible su un processore, è sufficiente sapere il WCET dei task e verificare che la somma all'interno di ogni slot sia minore o uguale al minor cycle.

Il **vantaggio** principale di timeline scheduling è la sua **semplicità**. Può essere implementato tra programmando un timer che interrompa con periodo uguale al minor cycle e scrivendo un ciclo main dove i task vengono chiamati nell'ordine dato dal major cycle, inserendo punti di sincronizzazione a ogni inizio del minor cycle. Dato che la sequenza di attivazione dei task non è determinata da un algoritmo di scheduling nel kernel, ma è determinata dalle chiamate effettuate dal main, non vi sono context switch, dunque l'overhead a runtime è molto basso. Nonostante questi vantaggi, vi sono dei problemi,

ad esempio l'alta fragilità in situazioni di overload. Se un task non termina nel limite del minor cycle, può essere continuato o abortito. In entrambi i casi, il sistema entra in uno stato critico. Se il task fallimentare rimane in esecuzione esso può causare un effetto a domino sugli altri task, rompendo l'intero schedule (timeline break). Se un task viene abortito mentre compie azioni su risorse condivise, esso può lasciare le risorse in uno stato inconsistente e causare un comportamento non corretto. Inoltre è particolarmente sensibile a cambiamenti nei requisiti.

6.3 Rate Monotonic Scheduling (RMS)

Funziona tramite una semplice regola che assegna la priorità a un task in base alla sua frequenza di attivazione. In particolare, task con periodi più brevi hanno priorità più alta. Dato che i periodi sono costanti, RM è un sistema fixed priority, ovvero la priorità P_i di un task i non cambia durante l'esecuzione. RM è intrinsecamente preemptive, ovvero un task con priorità più alta può interrompere un task con priorità più bassa. È stato dimostrato che RM è ottimale rispetto a tutti gli assegnamenti di priorità fissi, nel senso che nessun altro sistema di scheduling a priorità fissa può schedare un task set che non può essere schedato da RM.

È utile conoscere U_{lub} per RM su un task set di n task, che è dato da:

$$U_{lub} = n(2^{\frac{1}{n}} - 1) \quad (6)$$

n	U_{lub}	n	U_{lub}
1	1.000	6	0.73
2	0.828	7	0.729
3	0.780	8	0.724
4	0.757	9	0.721
5	0.743	10	0.718

(a)
(b)

Table 1: Valori di U_{lub} per RM al variare di n