

# Systèmes concurrents et distribués

Pierre Leone

[pierre.leone@unige.ch](mailto:pierre.leone@unige.ch)

Jérémie Martin

# Plan du cours

- Programmation concurrente
  - Exclusion mutuelle
  - Synchronisation
  - synchronisation wait-free
- Programmation distribuées
  - Sockets/RMI en Java
  - ...

# Plan du cours II

But:

- Considérer les problèmes classiques de programmation concurrente/distribuée
- Comprendre les contraintes liées aux applications concurrentes
- Illustrer les problèmes/solutions avec Java
- Discuter le langage Java le plus complètement possible

# Java

Pourquoi programmer en Java?

- C'est le seul langage pour lequel il existe un modèle formel de programmation
- Tout le monde parmi vous sait programmer en Java (séquentiel).

# Java pour illustrer

Le paquetage `java.util.concurrent` met à disposition des outils 'prêts à l'emploi'.

On va s'intéresser à l'implémentation de ces objets, le cours utilise Java pour illustrer les concepts, mais ne se limite pas à passer en revue les outils offerts par les librairies Java

# Références

*Concurrent and distributed computing in Java, Vijay K. Garg*

*Concurrent and Real-Time programming in Java, Andy Wellings*

*Java Concurrency in practice, Brian Goetz*

*The art of multiprocessor programming, Maurice Herlihy, Nir Shavit*

En français:

*Programmation concurrente et temps réel avec Java, Luigi Zaffalon*

# Introduction

Un programme **concurrent** est un programme qui peut s'exécuter en parallèle.

Un **processus** est un programme séquentiel qui compose un programme concurrent.

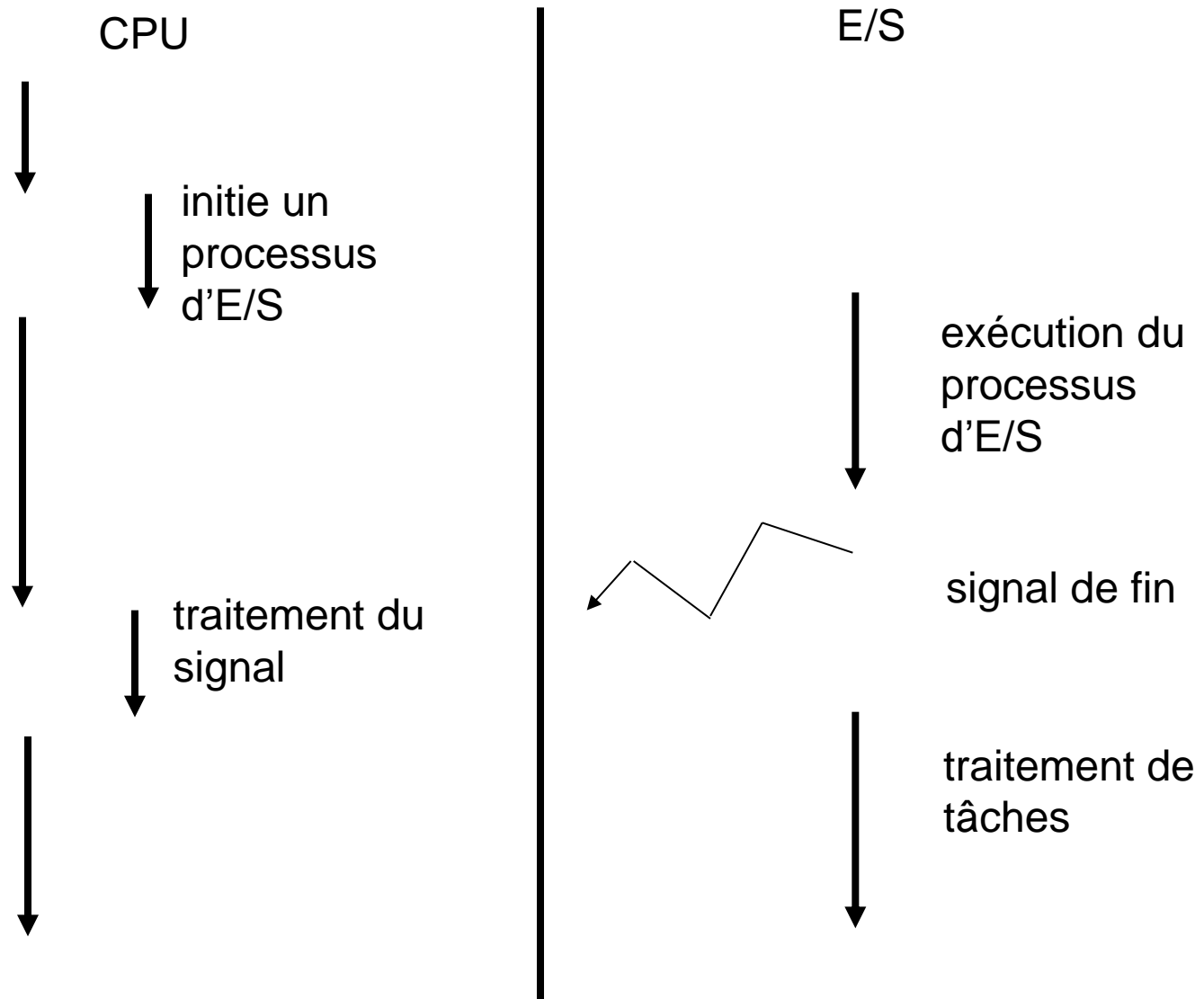
Des processus qui partagent une zone mémoire pour les données/programme sont appelés des **threads**.

# Motivation pour écrire un programme concurrent

- Utiliser le maximum des ressources d'un processeur:
  - par exemple un programme qui effectue des entrées/sorties intensivement et doit aussi effectuer des calculs.
- Permettre une implémentation réellement parallèle:
  - Un programme concurrent peut être exécuté par un système multi-processeur
- Pour modéliser/utiliser des interfaces qui sont réellement parallèles tels que des robots.



# Entrées/Sorties



# Critique des programmes concurrents

- Les mécanismes nécessaires à l'implémentation de la concurrence pénalisent l'exécution du programme exécuté par un système uni-processeur (overhead)

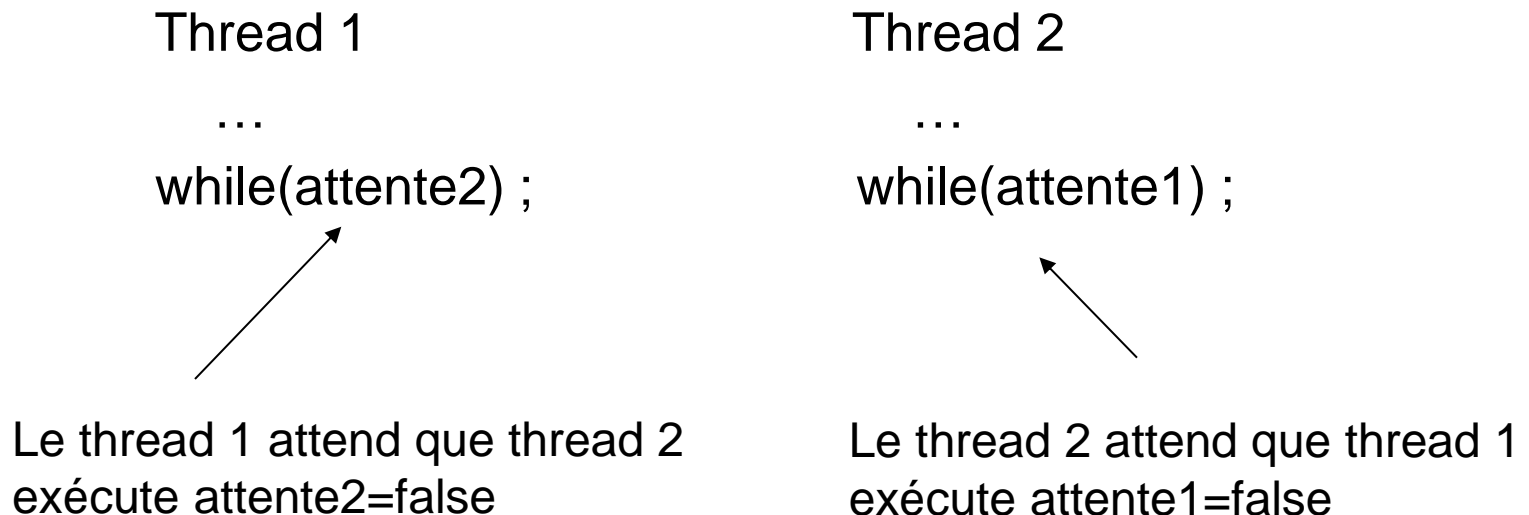
Il faut considérer la programmation concurrente comme un paradigme de programmation. Les simplifications apportées à la conception du programme l'emportent sur les questions d'efficacité (idem que langage assembleur – langage abstrait).

Nous permet de modéliser les systèmes réels, généralement concurrent.

# Problèmes classiques en programmation concurrente

Inter blocage (deadlock):

- Survient lorsque les entités d'un programme attendent une action d'une autre entité pour continuer



# Problèmes classiques en programmation concurrente

## Interférences:

- Plusieurs entités concurrentes modifient une donnée en même temps, la donnée peut finalement être corrompue

Par exemple la donnée est une date de naissance jj/mm/aaaa

Thread1 exécute jj = 1, mm = 12, aaaa = 1982

Thread2 exécute jj = 20, mm = 1, aaaa = 1946

Après exécution le résultat est jj = 1, mm = 1, aaaa=1946

# Problèmes classiques en programmation concurrente

Insuffisance de ressources (starvation):

- Une entité n'arrive pas accéder une ressource, cette dernière étant perpétuellement utilisée par d'autres entités.

# Propriétés désirées

## Sûreté (safety):

- un événement indésirable ne doit pas se produire pendant l'exécution, en particulier pas d'interférences entre les différentes entités.

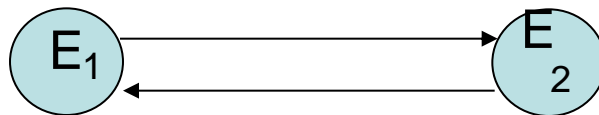
## Vivacité (*liveness*) :

- un événement souhaité arrivera nécessairement, en d'autres termes l'exécution du programme ne résultent pas en interblocages ou insuffisance de ressources

# Inter blocage

Un programme doit satisfaire quatre conditions nécessaires pour qu'un inter blocage soit possible

- **Exclusion mutuelle:** Une ressource non partageable simultanément doit être accédée, une telle ressource doit être partagée séquentiellement.
- **Hold and wait:** Les entités impliquées dans l'inter blocage doivent posséder l'accès à une ressource non partageable simultanément et attendre l'accès à une autre ressource non partageable.
- **Pas de préemption:** Les ressources doivent impérativement être explicitement libérées par l'entité qui dispose de l'accès.
- **Attente circulaire:** Les entités forment une chaîne dans laquelle chacune dispose de l'accès à une ressource et attend pour accéder la ressource dont dispose la prochaine entité dans la chaîne.



# Les solutions

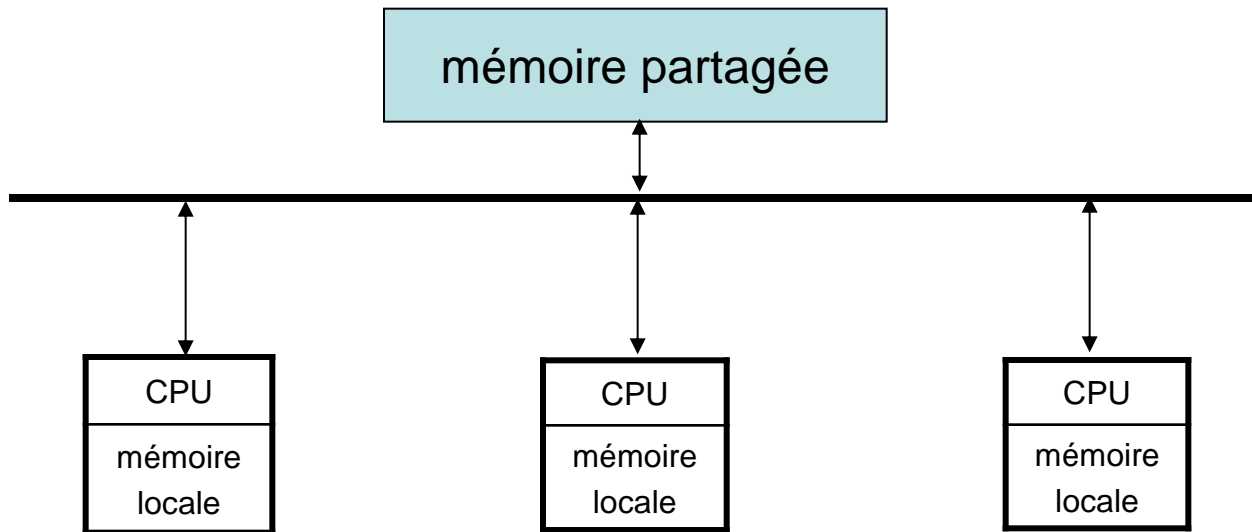
Il existe trois solutions pour prévenir les inter blocages

- S'assurer qu'au moins une des quatre conditions nécessaires ne peut pas se produire. Par exemple en interdisant une entité d'attendre l'accès à une ressource si elle est en possède déjà une (**deadlock avoidance**).
- Utiliser un algorithme d'allocation des ressources qui gère dynamiquement les accès en évitant les inter blocages (**deadlock avoidance algorithm**).
- Utiliser un système de détection des inter blocages qui effectue les actions nécessaires au déblocage. (Par exemple en retirant l'accès à un processus et en donnant l'accès à un autre par préemption).



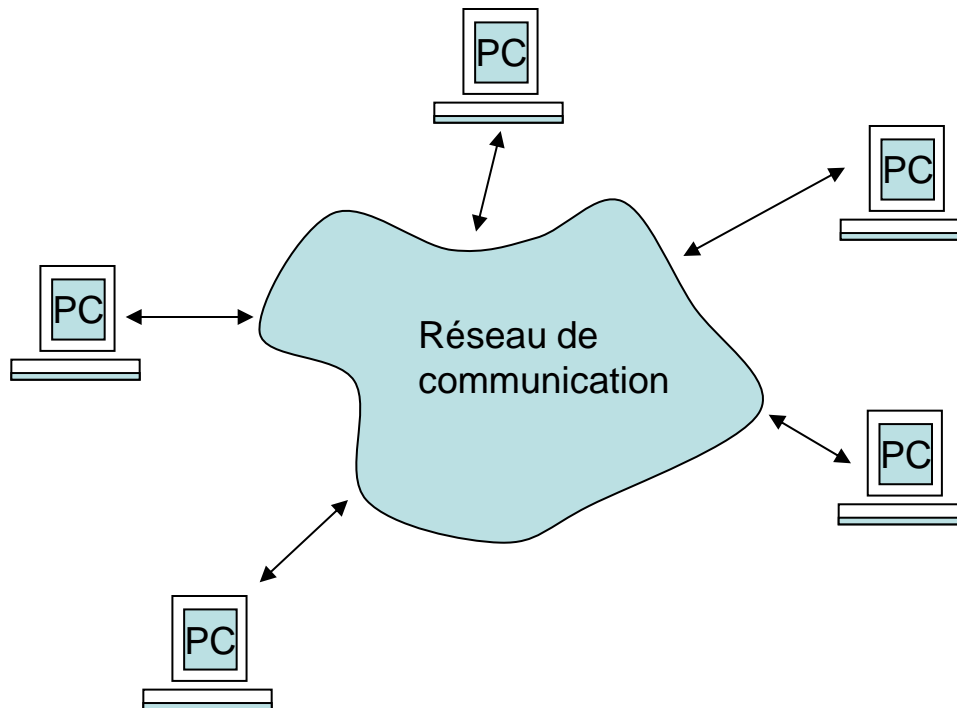
# Généralités

On réserve généralement le terme **programmation parallèle** aux programmes qui s'exécutent vraiment en parallèle et communiquent en utilisant une mémoire partagée. Un programme concurrent n'est pas nécessairement parallèle (systèmes multitâches).



# Généralités

Un **système distribué** est composé de systèmes qui communiquent entre eux par un réseau de communication en utilisant des **messages**.



# Parallèles / distribués

Pourquoi distinguer les systèmes parallèles et distribués. En effet, on peut simuler le passage de messages avec un système parallèle et simuler une mémoire partagée avec un système distribué.

Néanmoins, chacun des systèmes physiques possède des propriétés qui lui sont propres

# Parallèles / distribués

- **Extensibilité (scalability):** Les systèmes distribués sont plus facilement extensible que les systèmes parallèles pour lesquels les accès à la mémoire commune deviennent critiques si le nombre de CPU est trop important
- **Modularité/heterogeneité:** il est facile d'ajouter/enlever des éléments d'un système distribué et ces éléments ne doivent pas être identiques.
- **Partage des données/ressources:** les systèmes distribués sont basés sur le partage de données/ressources entre différents systèmes

# Parallèles / distribués

- **Structure géographique:** Certaines applications sont géographiquement distribuées et il est préférable d'effectuer le plus possible de calcul localement pour limiter la quantité de données transférées (réseaux sans fil)
- **Fiabilité:** Les systèmes distribués sont moins susceptibles d'être affectés par une machine en panne
- **Efficacité:** Les systèmes parallèles sont généralement utilisés pour leur efficacité.

# Caractéristiques des systèmes

- Les systèmes distribués ne partagent pas de données dans une mémoire partagée. Pour observer l'état du système il faut développer des algorithmes particuliers.
- Les systèmes distribués ne disposent pas d'horloges communes permettant au différents processus de se synchroniser, il faut utiliser d'autres concepts pour raisonner sur l'évolution des programmes (notion de causalité)
- un système distribué est asynchrone s'il n'existe pas de borne supérieure pour le temps de communication entre deux processus. Dans cette situation il est difficile de distinguer un système lent d'un système ne fonctionnant plus.

# Contraintes de développement

- **Tolérance aux pannes:** Le système global doit continuer de fonctionner si certains sous-systèmes sont défectueux.
- **Transparence:**
  - Les différents modes de codages des données (little/big endian) doivent être transparents à l'utilisateur (access transparency)
  - Lorsqu'un utilisateur utilise une ressource partagée il ne doit pas se soucier
    - de localiser la ressource (location transparency)
    - si la donnée est dupliquée (replication transparency)
    - si la donnée est partagée (concurrency transparency)

# Contrainte de développement

- **Flexibilité:** Les systèmes doivent permettre l'interaction de systèmes différents, par exemple en utilisant des interfaces de communication normalisées
- **Extensibilité (scalability):** Les systèmes doivent prendre en compte l'évolution du nombre d'utilisateurs ou des ressources. Par exemple, un système centré sur un unique serveur ne fonctionnera pas correctement si le nombre d'utilisateurs devient trop important.



# Programmation concurrente en Java

Le langage java permet de composer un programme en différents **threads**, qui sont exécutés par la même machine virtuelle java (JVM) et donc peuvent partager les mêmes ressources.

Le modèle concurrent de java est basé sur la notion d'objets actifs qui s'exécutent concurremment. Ces objets encapsulent un thread. Pour le programmeur cela revient à utiliser des instances de la classe Thread.

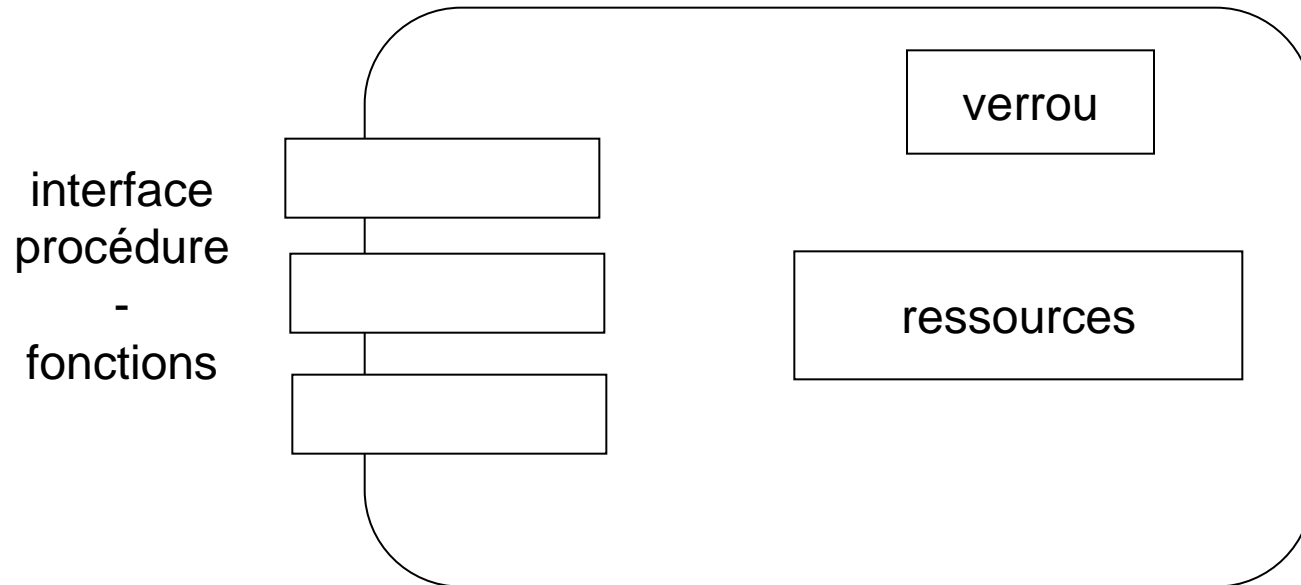
# Programmation concurrente en java

Les mécanismes de communication et de synchronisation proposés par Java sont inspirés de la notion de **moniteur**.

Un moniteur encapsule une/des ressources partagées (généralement des variables) et propose des procédures/fonctions qui permettent de manipuler ces ressources.

Les moniteurs sont des extensions des objets car les procédures/fonctions s'exécutent de manière **atomique**, c'est-à-dire que les différentes exécutions ne peuvent pas interférer. En effet, les exécutions sont **mutuellement exclusives**: A chaque moniteur est associé un verrou (lock) que chaque thread/processus doit acquérir avant de d'exécuter une procédure/fonction et restituer après l'exécution.

# Un moniteur

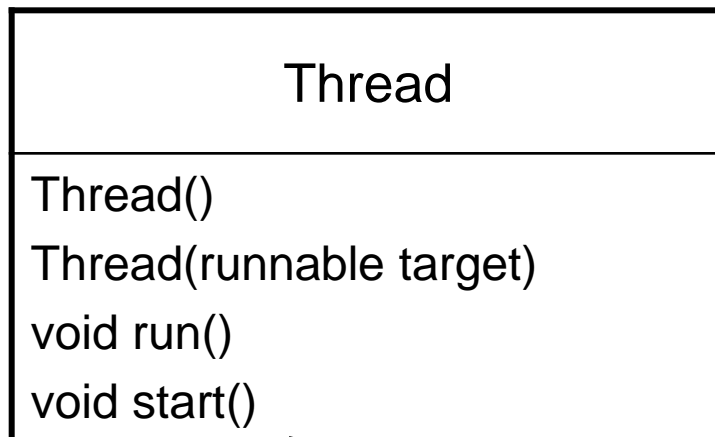


L'exclusion mutuelle est assurée pour l'accès aux ressources encapsulées, une liste des threads en attente est aussi associées au moniteur.

# Hello World

En Java un thread est une instance d'une classe.

On peut définir une telle classe en la dérivant de la classe Thread (java.lang.Thread).



effectue les opérations nécessaires pour que le thread soit pris en compte par la JVM, puis exécute la méthode run()

# Hello world

```
public class HelloWorldThread extends Thread
```

```
{
```

```
    public void run()
```

```
{
```

```
        System.out.println("Hello, world");
```

```
}
```

point d'entrée du  
programme



```
    public static void main(String[] args)
```

```
{
```

```
        HelloWorldThread t = new HelloWorldThread();
```

```
        t.start();
```

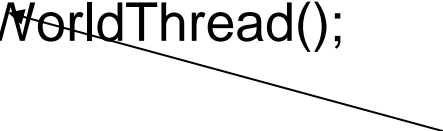
```
}
```

```
}
```

exécution du thread



création d'un objet de  
la classe



# plusieurs threads

```
class Ecrit extends Thread
```

```
{  
    public Ecrit(String texts, int nb)  
    { this.texte = texts;  
      this.nb = nb;  
    }  
    public void run()  
    { for(int i=0; i<nb; i++)  
      System.out.print(texte);  
    }  
    private String texte;  
    private int nb;  
}
```

} variables locales, une copie par thread

# Plusieurs threads

```
public class TstThread1
{ public static void main(String args[ ])
  { Ecrit e1= new Ecrit("bonjour ", 10);
    Ecrit e2 = new Ecrit("bonsoir ", 12);
    Ecrit e3 = new Ecrit("\n ",5);

    e1.start();
    e2.start();
    e3.start();
  }
}
```

point d'entrée du programme

création des objets

exécution des threads

# Premier exemple d'exécution

```
K:\Cours\systemesdistribues\prgJava\chap2>
K:\Cours\systemesdistribues\prgJava\chap2>
K:\Cours\systemesdistribues\prgJava\chap2>
K:\Cours\systemesdistribues\prgJava\chap2>
K:\Cours\systemesdistribues\prgJava\chap2>javac Ecrit.java
K:\Cours\systemesdistribues\prgJava\chap2>javac TstThread1.java
K:\Cours\systemesdistribues\prgJava\chap2>java TstThread1
bonjour bonjour bonjour bonjour bonjour bonjour bonjour bonjour bonjour bonjour
bonsoir bonsoir bonsoir bonsoir bonsoir bonsoir bonsoir bonsoir bonsoir bonsoir
bonsoir bonsoir
K:\Cours\systemesdistribues\prgJava\chap2>
```



# Premier exemple d'exécution

On constate que les threads se sont exécutés dans l'ordre dans lequel ils ont été créés. Les threads sont toujours exécutés dans le même ordre (Pas une règle: dépendent de l'environnement).

Un thread peut explicitement interrompre son exécution pendant une période de temps  $t$  (millisecondes) en exécutant la méthode *sleep(t)*.

```
public void static mySleep(int time)
{ try {
    Thread.sleep( time );
} catch{ InterruptedException e) {} // le bloc try est imposé par sleep(t)
}
```

# Utilisation de Thread.sleep(t)

```
public void run ()  
{ try  
  { for( int i=0 ; i<nb ; i++)  
    { System.out.print(texte);  
      sleep(attente); // on étend la classe Thread  
    }  
  } catch (InterruptedException e) {} // nécessaire pour sleep()  
}
```

# Attente aléatoire

```
import java.util.Random;

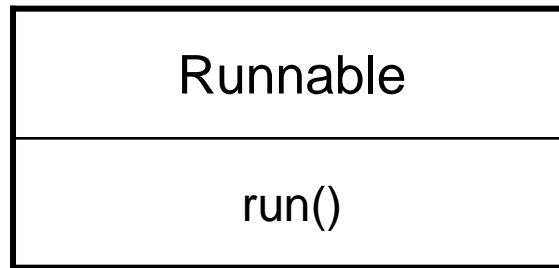
try {
    Thread.sleep(r.nextInt(10));
} catch (InterruptedException ie) {}
```

# Remarques

- La méthode *main()* qui correspond au programme principal est le **thread principal**.
- La méthode *sleep()* permet de donner la main à un autre thread (y compris le thread principal).
- On peut invoquer seulement la méthode *run()*, le programme s'exécute mais dans un seul thread.
- La méthode *start()* ne peut être appelée qu'une seule fois pour un objet donné.

# Interface Runnable

Il est possible de créer un thread en implémentant l'interface *Runnable*, laquelle comporte une seule méthode *run()*.



# Interface Runnable et classe Thread

En fait, la classe *Thread* implémente l'interface *Runnable*

```
package java.lang;
```

```
public class Thread extends Object implements Runnable
{ public Thread();
  public Thread(String name);
  public Thread(Runnable target);
  public Thread(Runnable target, String name);
  public Thread(Runnale target, String name, long stackSize);

  public static Thread currentThread();
  public void run();
  public void start();
      .....
}
```

# Exemple

**class** Ecrit **implements** Runnable

{ **public** Ecrit (String texte, **int** nb, **long** attente)

{ this.texte = texte;

  this.nb = nb;

  this.attente = attente;

}

**public void** run()

{ **try**

{ **for** (int i=0; i<nb; i++)

{ system.out.print(texte);

  Thread.sleep(attente);

} } **catch** (InterruptedException e) {}

}

**private** String texte; **private** int nb; **private long** attente;

}

}  
définition de la  
méthode run()

# Exemple (suite)

On peut maintenant créer une instance de la classe Ecrit

```
Ecrit e1 = new Ecrit("bonjour ", 10,5);
```

Ensuite on crée le thread

```
Thread t1 = new Thread(e1);
```

Que l'on peut exécuter

```
t1.start();
```



# Définition de la méthode *start()*

Dans la classe `Ecrit`, il est possible de définir la méthode `start()` comme dans la classe `Thread`

```
public void start()  
{ Thread t = new thread(this);  
  t.start();  
}
```

# Interruption d'un thread

Java dispose d'un mécanisme permettant à un thread d'en interrompre un autre.

La méthode `Thread.interrupt()` positionne un indicateur qui indique au thread qu'une requête 'interrupt' à été déposée.

Un thread peut connaître l'état de l'indicateur à l'aide de la méthode statique *`interrupted()`*.

Thread 1

`Thread2.interrupt();`

Thread 2

....

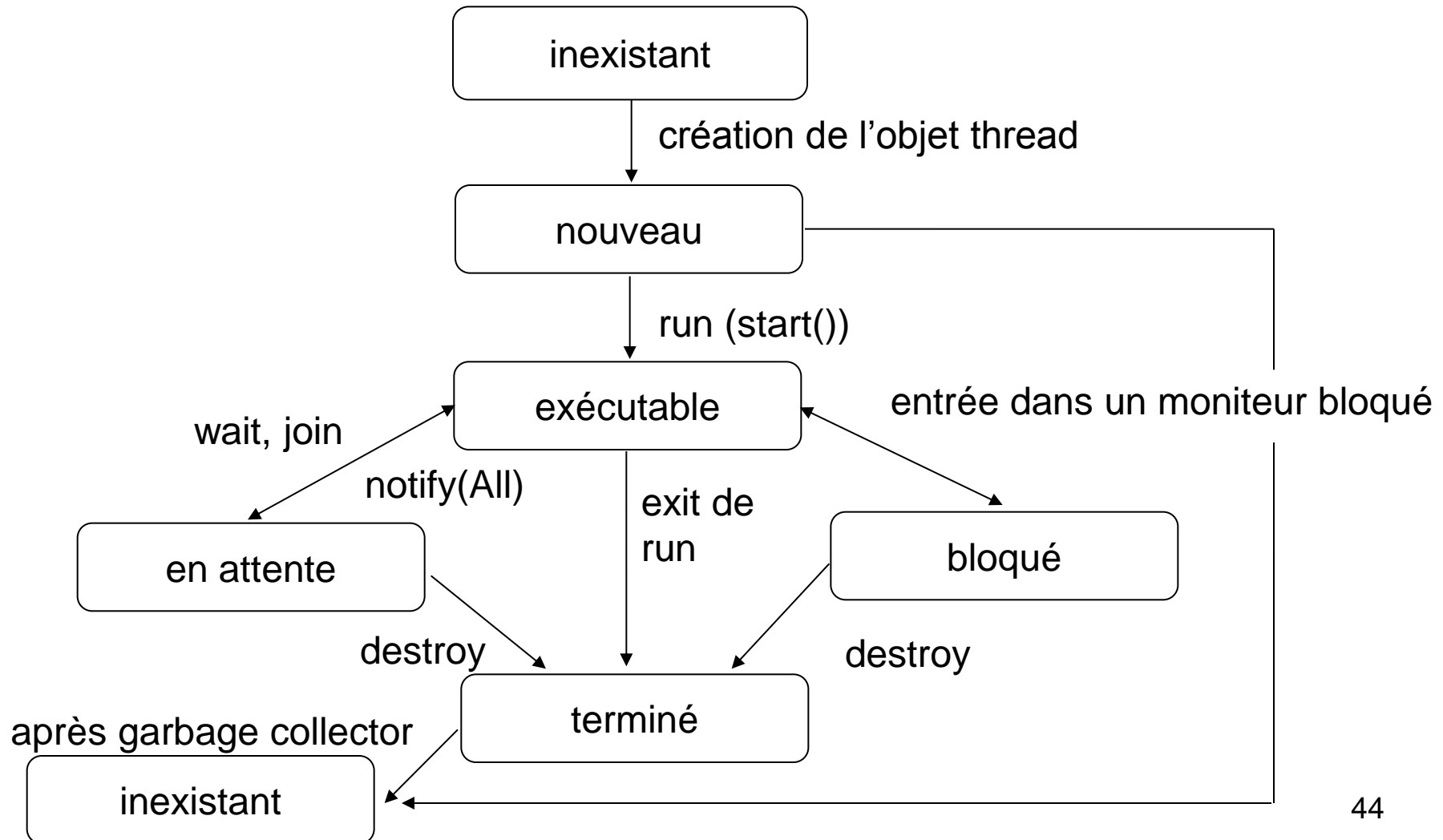
**`if(interrupted())`**

**`return;`**      *// fin du thread*

# Remarques

- La prise en compte du signal 'interrupt' par le thread reste sous sa propre responsabilité.
- La méthode `Tread.interrupt()` n'a pas d'effet immédiat sur le thread.
- Ce dernier doit périodiquement scruter l'indicateur en exécutant la méthode *interrupted()*.
- La méthode *interrupted()* positionne l'indicateur à *false* après lecture.
- La méthode *isInterrupted()* s'exécute comme la méthode *interrupted()* mais ne modifie pas l'indicateur après lecture.
- Les méthodes *wait()*, *sleep()*, *join()* testent l'état de l'indicateur et lèvent l'exception *InterruptedException* si l'indicateur est positionné à *true*.

# Les différents états d'un thread



# Ordonnancement des threads

A chaque thread est associée une priorité qui peut être  
MAX\_PRIORITY, MIN\_PRIORITY, NORM\_PRIORITY (par défaut)

Les méthodes de la classe Thread

```
public final void setPriority(int newPriority)
```

```
public final int getPriority()
```

permettent de modifier/lire la priorité d'un thread.

Lorsque l'ordonnanceur peut donner la main (p. ex. fin du time slice) à un thread il choisit celui de plus haute priorité, si plusieurs threads sont candidats le choix dépend de la JVM.

Si un thread plus prioritaire que le courant devient exécutable, l'ordonnanceur lui donne la main, le courant passant dans l'état exécutable.

# Un exemple avec *join()*

La méthode *join()* permet à un thread d'attendre qu'un autre thread ait fini son exécution.

Pour illustrer comment on peut utiliser *join()*, on considère le calcul des nombre de Fibonacci  $F_n$   $n \geq 0$ . Ils sont définis récursivement par:

$$F_0 = F_1 = 1$$

$$F_n = F_{n-1} + F_{n-2} \quad n \geq 2$$

L'idée est de procéder comme on le ferait avec un programme récursif. C'est-à-dire, pour calculer  $F_n$  on exécute deux threads, l'un qui calcule  $F_{n-1}$  et l'autre  $F_{n-2}$ . Une fois que les deux threads ont terminés leur exécution, on calcule .

# Fibonacci

```
Public class Fibonacci extends Threads {  
    int n; int result;  
    public Fibonacci( int n) {  
        this.n = n;  
    }  
    public void run() {  
        if ((n==0) || (n==1)) result = 1;  
        else {  
            Fibonacci f1 = new Fibonacci(n-1);  
            Fibonacci f2 = new Fibonacci(n-2);  
            f1.start();  
            f2.start();  
            try{  
                f1.join();  
                f2.join();  
            } catch (InterruptedException e) {};  
            result = f1.getResult() + f2.getResult();  
        }  
    }  
}
```

# Fibonacci (suite)

```
public int getResult() {  
    return result;  
}
```

```
public static void main(String[] args) {  
    Fibonacci f1 = new Fibonacci(Integer.parseInt(args[0]));  
    f1.start();  
    try{  
        f1.join();  
    } catch (InterruptedException e) {};  
    System.out.println(" La réponse est " + f1.getResult());  
}  
}
```



# Exclusion mutuelle

# Problème de l'exclusion mutuelle

Lorsque plusieurs threads accèdent des variables partagées, les accès à ces variables doivent être synchronisés.

Thread 1

```
....  
x = x + 1  
....
```

Thread 2

```
....  
x = x + 1  
....
```

x est une variable partagée

# Exclusion mutuelle

Les instructions  $x = x + 1$  se décomposent en plusieurs sous-instructions:

1. lecture de  $x$
2. addition ( $x+1$ )
3. écriture de  $x$

il existe un scénario qui ne modifie pas la variable  $x$  deux fois

Thread 1

lecture  $x$

addition

écriture de  $x$

Thread 2

lecture de  $x$

addition

écriture de  $x$

# Atomicité

Cet exemple d'exécution montre que la variable  $x$  ne peut pas être utilisée comme compteur par exemple.

Pour résoudre le problème l'instruction  $x=x+1$  doit s'exécuter de manière **atomique**, c'est-à-dire que les deux accès à la variable  $x$  (l/e) doivent s'effectuer de manière **indivisible**.

Une portion de code qui doit s'exécuter de manière atomique est une **section critique (SC)**.

Une solution consiste à inclure l'instruction dans une section critique qui est protégée par un verrou.

# Section critique

On protège donc les instructions  $x=x+1$  par un protocole d'entrée/sortie de la section critique

Thread 1

Entrée en SC

$x = x + 1$

Libère la SC

Thread 2

Entrée en SC

$x = x + 1$

Libère la SC

# verrou

un verrou peut-être vu comme un objet Java dont l'interface est:

```
public interface Lock      // voir l'interface Runnable pour l'utilisation
{
    public void requestCS(int pid); // requête pour entrer en SC
    public void releaseCS(int pid); // indique que le thread quitte la SC
}
```

La méthode *requestCS(int pid)* est bloquante, c'est-à-dire que si un processus se trouve en section critique au moment de son exécution le processus appelant est bloqué.

Les méthodes doivent assurer que jamais plus d'un processus se trouve en SC (safety, sûreté) et qu'un processus qui désire entrer en SC le fera (liveness, vivacité).

# Verrou

## Thread1

```
...  
requestCS(1);  
...  
code de la section critique  
...  
releaseCS(1);
```

} Un thread qui effectue requestCS avant que Thread1 ait exécuté releaseCS sera bloqué.

# Exceptions

Lorsqu'un thread à obtenu un verrou, il est important pour le bon fonctionnement du programme qu'il le libère quoi qu'il arrive, en particulier si une exception est levée pendant l'exécution du code correspondant à la section critique. Pour cela, on écrit:

```
mutex.requestCS();  
try {  
    ... le corps de la section critique ...  
} finally {  
    mutex.releaseCS(); // est exécutée qu'une exception soit levée  
}                      // ou pas...
```



# Programme test

```
import java.util.Random;

public class MyThread extends Thread {
    int myId; Lock lock; Random r= new Random();

    public MyThread(int id, Lock lock) {
        myId = id;           // chaque processus possède une identité propre
        this.lock = lock;    // les processus utilisent le même verrou pour accéder la SC
    }

    void nonCriticalSection() {
        System.out.println(myId + " n'est pas en SC ");
        mySleep(r.nextInt(1000));
    }

    void CriticalSection() {
        System.out.println(myId + " est en SC ");
        mySleep(r.nextInt(1000));
    }
}
```

# Programme test

```
public void run() {  
    while(true) {  
        lock.requestCS(myId);  
        // section critique  
        lock.releaseCS(myId);  
        // section non critique  
    }  
}  
  
public static void main(String [] args) throws Exception {  
    MyThreadt [];  
    int N = Integer.parseInt(args[0]);  
    t = new myThread[n];  
    Lock = lock new .....(N); compléter avec un algorithme mutex  
    for(int i=0; i<N; i++){  
        t[i]=new MyThread(i,lock);  
        t[i].start();  
    }  
}
```

# Première tentative – 2 processus

```
class Attempt1 implements Lock {  
    private boolean openDoor = true;  
    public void requestCS(int i) {  
        while (!openDoor) ; // attente active  
        openDoor = false; // verouille l'accès à la SC  
    }  
  
    public void releaseCS(int i) {  
        openDoor = true; // libère l'accès à la SC  
    }  
}
```

# Première tentative

Cette première tentative n'est pas correcte car si le processus perd la main après qu'il ait testé `openDoor = true` et avant qu'il ait exécuté `openDoor = false`, un processus peut entrer en section critique et finalement les deux processus se retrouveront simultanément en SC.

Le problème ici est que le test de la valeur de `openDoor` (lecture) et sa modification (écriture) ne sont pas exécutés de manière atomique.

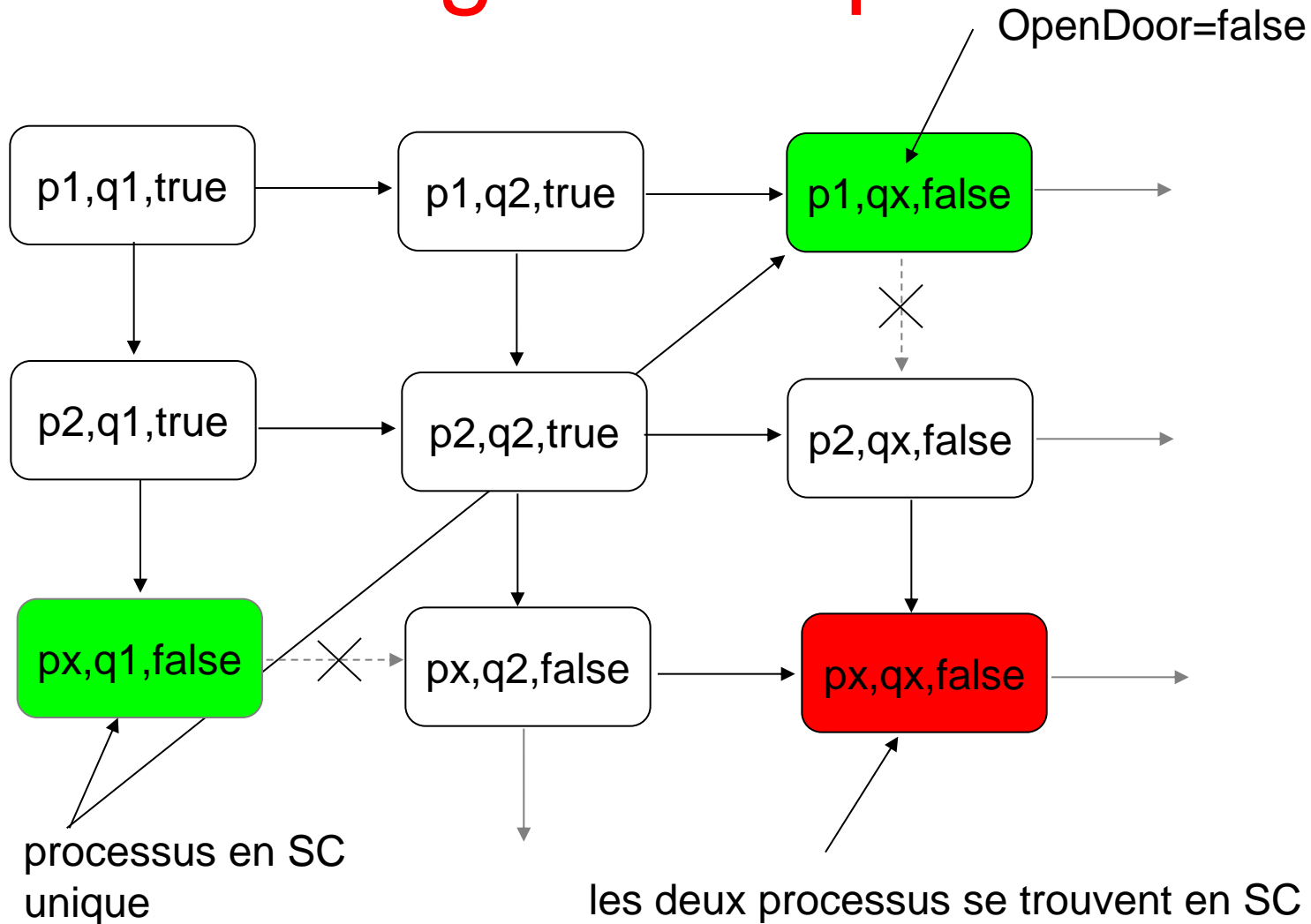
# Diagramme

```
public void requestCS(int i) {  
1    while (!openDoor) ;  
2    openDoor = false;  
}  
  
public void releaseCS(int i) {  
3    openDoor = true; // libère l'accès à la SC  
}
```

On numérote les lignes à exécuter et on utilise les lettres p et q pour désigner les deux processus. C'est-à-dire p2 indique que la prochaine instruction à exécuter par le processus p est celle qui se trouve en ligne 2.

L'état du programme est défini par le 3-tuple  $(p_i, q_j, \text{openDoor})$

# Diagramme partiel



# Deuxième tentative – 2 processus

On essaye de résoudre le problème en introduisant deux nouvelles variables partagées (une par processus) wantCS[0] et wantCS[1] que les processus utilisent pour signifier qu'ils désirent entrer en SC.

```
class Attempt2 implements Lock {  
    private boolean wantCS[] = {false, false};  
    public void requestCS(int i) {  
        wantCS[i] = true;    // réserve l'accès à la SC  
        while (wantCS[1-i]); // attente active si le second processus en SC  
    }  
    public void releaseCS(int i) {  
        wantCS[i] = false;  
    }  
}
```

# Deuxième tentative

Cette deuxième tentative assure bien **l'exclusion mutuelle**. En effet, si les deux processus se trouvent en section critique simultanément alors on a  $\text{wantCS}[0]=\text{wantCS}[1]=\text{true}$ .

Pour que le thread  $i$  entre en SC il faut qu'il teste  $\text{wantCS}[1-i]=\text{false}$  sinon il est bloqué dans la boucle **while**, c'est-à-dire que le thread  $1-i$  n'ait pas encore exécuté  $\text{wantCS}[1-i]=\text{true}$ . Ce processus va donc être bloqué par le test dans la boucle **while** (car  $\text{wantCS}[i]=\text{true}$ ).

On suppose implicitement que l'exécution de  $\text{wantCS}[i]=\text{true}$  par le processus  $i$  est vue immédiatement par le processus  $1-i$  (entrelacement des instructions).



# Deuxième tentative

Le problème avec cette deuxième tentative est que les deux processus peuvent positionner  $\text{wantCS}[i] = \text{wantCS}[1-i] = \text{true}$  et se trouvent simultanément bloqués par la boucle **while**.

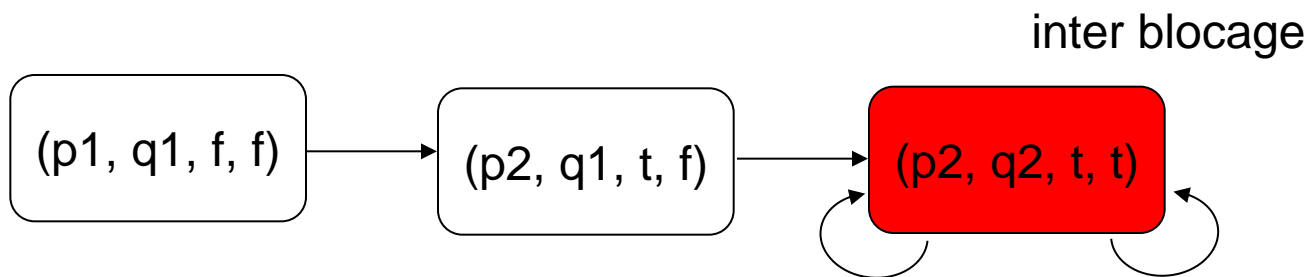
On a possibilité d'inter blocage.

Pour dessiner le diagramme, l'état de l'algorithme est déterminé par le processus actif  $p$  ( $i=0$ ) et  $q$  ( $i=1$ ), l'état  $\text{wantp}$  et  $\text{want q}$

# Diagramme

```
class Attempt2 implements Lock {  
    boolean wantCS[] = {false, false};  
    public void requestCS(int i) {  
1        wantCS[i] = true;    // réserve l'accès à la SC  
2        while (wantCS[1-i]); // attente active si le second processus en SC  
    }  
    public void releaseCS(int i) {  
3        wantCS[i] = false;  
    }  
}
```

L'algorithme peut donner lieu à un inter blocage.



# Troisième tentative

Pour éviter l'inter blocage on peut " inverser" la stratégie et utiliser une variable booléenne pour favoriser l'accès à l'autre processus.

```
class Attempt3 implements Lock {  
    private int turn = 0;  
    public void requestCS(int i) {  
        while(turn==1-i); // attente active, ce n'est pas notre tour  
    }  
  
    public void releaseCS(int i) {  
        turn = 1-i; // on sort de SC et on libère pour l'autre processus  
    }  
}
```

# Troisième tentative

Avec cette solution les processus sont obligés d'alterner leurs accès en section critique. En effet, si le processus  $i$  quitte la section critique il peut y retourner seulement si le processus  $1-i$  y accède et lui permet l'accès.

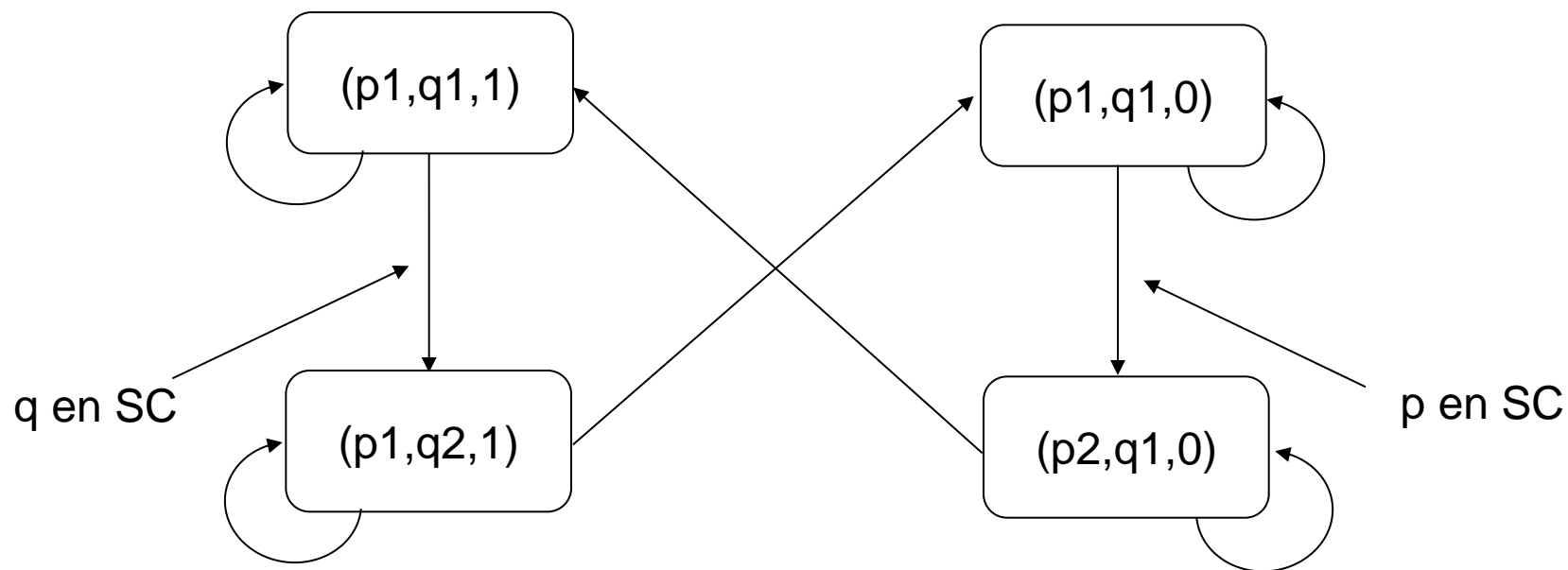
Si le processus  $1-i$  n'accède pas la section critique, alors le processus  $i$  ne peut plus jamais l'accéder.

# Diagramme

```
class Attempt3 implements Lock {  
    private int turn = 0;  
    public void requestCS(int i) {  
1        while(turn==1-i); // attente active, ce n'est pas notre tour  
    }  
  
    public void releaseCS(int i) {  
2        turn = 1-i; // on sort de SC et on libère pour l'autre processus  
    }
```

On a deux processus p ( $i=0$ ) et q ( $i=1$ ) et une variable partagée turn, (px,py,turn) définit l'état courant de l'algorithme

# Diagramme



L'algorithme assure l'exclusion mutuelle, l'état (p2, q2, ...) n'apparaît pas dans le diagramme.

# Algorithme de Peterson

*G.L. Peterson Myths about the mutual exclusion problem, Information processing letters, 12(3):115-116, 1981.*

```
class PetersonAlgoritm implements Lock {  
    private boolean wantCS[] = {false, false};  
    private int turn = 1;  
    public void requestCS(int i) {  
        int j = 1-i;  
        wantCS[i] = true;  
        turn = j;  
        while (wantCS[j] && (turn == j)); // les deux processus ne peuvent plus  
    }                                     // se trouver simultanément en SC,  
    public void releaseCS(int i) {      // l'alternance n'est plus nécessaire  
        wantCS[i] = false;  
    }  
}
```

écriture de wantCS  
écriture de turn  
lecture de wantCS et turn

# Algorithme de Peterson

L'algorithme assure l'exclusion mutuelle:

On suppose le processus p ( $i=0$ ) se trouve en SC, c'est-à-dire p doit lire  $wantCS[1]=false$  ou  $turn=0$

**Cas 1:** p lit  $wantCS[1]=false$ .

avant de rentrer en SC le processus q ( $i=1$ ) doit positionner  $wantCS[1]=true$ .

p lit  $wantCS[1]=false \longrightarrow$  q écrit  $wantCS[1]=true$  **hypothèse**

la flèche indique une relation de précédence.

q écrit  $wantCS[1]=true \longrightarrow$  q écrit  $turn=0$  ordre du programme (p.o)  
(intra-processus)



# Algorithme de Peterson

On obtient la séquence suivante:

p écrit turn=1  $\xrightarrow{\text{p.o.}}$  p lit wantCS[1]=false  $\xrightarrow{\text{hyp.}}$  q écrit wantCS[1]=true  
 $\xrightarrow{\text{p.o.}}$  q écrit turn=0  $\xrightarrow{\text{p.o.}}$  q lit turn

**alors q lit turn = 0**

p écrit wantCS[0]=true  $\xrightarrow{\text{p.o.}}$  p lit wantCS[1]=false  $\xrightarrow{\text{par hypothèse}}$   
q écrit wantCS[1]=true  $\longrightarrow$  q lit wantCS[0]

**alors q lit wantCS[0]=true**

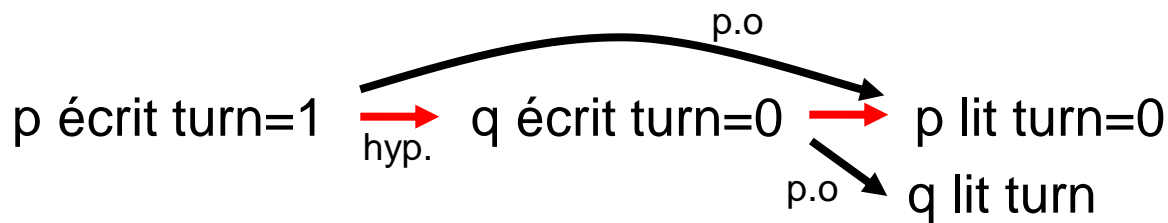
**le processus q est bloqué par la boucle while**

q ne peut pas entrer en SC

# Algorithme de Peterson

**Cas 2:** p lit turn=0

Dans ce cas, q doit positionner turn=0 après que p exécute turn=1 et avant que p lise turn



**q doit nécessairement lire turn=0** (première condition d'attente)

On insiste sur le fait qu'en reconstruisant la séquence des événements on doit être cohérent avec la séquence induite par l'ordre du programmes (p.o.)

# Algorithme de Peterson

Le processus p écrit `wantCS[0]=true` avant d'écrire `turn=1`

p écrit `wantCS[0]=true`  $\xrightarrow{\text{p.o.}}$  p écrit `turn=1`  $\xrightarrow{\text{hyp.}}$  q écrit `turn=0`  $\xrightarrow{\text{p.o.}}$   
q lit `wantCS[0]`

**q doit nécessairement lire `wantCS[0]=true`** (deuxième condition d'attente)

finalement q ne peut pas entrer en SC.

# Algorithme de Peterson

Aucune exécution du protocole ne peut générer un **inter-blocage**.

En effet, la condition pour que les deux processus soient en attente simultanément est que:

$(\text{wantCS}[1]=\text{true} \ \&\& \ \text{turn}=1) \ \&\& \ (\text{wantCS}[0]=\text{true} \ \&\& \ \text{turn}=0)$



De plus, un processus accède toujours à la section critique (**insuffisance de ressources, starvation**). En effet, supposons p en SC et q bloqué. Lorsque p quitte la section critique il positionne  $\text{wantCS}[0]=\text{false}$ .

Si p désire entrer en SC avant que q ait remarqué que  $\text{wantCS}[0]=\text{false}$ , il exécute  $\text{wantCS}[0]=\text{true}$  (bloquant) et  $\text{turn}=1$  qui donne l'accès au processus q.

# Algorithme de Lamport

Le nom original de l'algorithme est : Lamport's bakery algorithm car l'idée est que chaque processus qui exécute le protocole d'entrée en SC se voit attribuer un numéro et le processus qui possède le numéro le plus petit se voit attribuer l'accès en SC

Une difficulté est qu'on ne peut pas assurer que chaque processus reçoit un numéro unique (les processus sont identifiés).

Le protocole d'entrée est divisé en deux:

1. Chaque processus choisit un numéro plus grand que les numéros déjà attribués.
2. Chaque processus teste s'il peut entrer en SC
  - a) S'assure qu'aucun processus est en phase 1.
  - b) S'assure qu'il possède le plus petit numéro, en cas d'égalité les identificateurs de processus font la différence

# Lamport's Bakery algorithm

*L. Lamport, A new solution of Dijkstra's concurrent programming problem, Comm. of the ACM, 17(7), 1974.*

**class Bakery implements Lock {**

**int** N;

**volatile boolean** [] choosing; // processus en phase 1

**volatile int** [] number; // gestion de la file d'attente

**public Bakery(int numProc) {**

N = numProc;

choosing = **new int** [N];

**for** (**int** j = 0; j < N; j++) {

choosing[j] = **false**;

number[j] = 0;

}

}

initialisation du verrou  
pas de processus en phase 1.  
numéro 0 attribué à tous les processus

# Lamport's bakery algorithm

```
public requestCS(int i) {  
    choosing[i] = true;  
    for (int j = 0; j < N; j++)  
        if (number[i] < number[j])  
            number[i] = number[j];  
    number[i]++;    // on choisi le plus grand numéro  
    choosing[i]=false  
    for (int j = 0; j < N; j++) {  
        while (choosing[j]); // attente proc. en phase 1.  
        while ((number[j] != 0) && ((number[j] < number[i]) ||  
            ((number[j]==number[i] && j < i))) ; // attente active  
    }  
}
```

Phase 1. choix du numéro

$(\text{number}[j], j) < (\text{number}[i], i)$

# Lamport's bakery algorithm

```
public void releaseCS(int i) { // protocole de sortie de SC
    number[i] = 0;
}
```

La relation d'ordre introduite est:

$(\text{number}[i], i) < (\text{number}[j], j)$  si  $\text{number}[i] < \text{number}[j]$   
ou  $(\text{number}[i] == \text{number}[j] \ \&\& \ i < j)$

c'est une relation d'ordre totale si on associe a chaque processus un numéro différent.



# Preuve de l'algorithme

## 1<sup>ère</sup> assertion:

Si un processus  $P_i$  se trouve en SC et un autre processus  $P_k$  a déjà choisi un numéro alors

$$(number[i], i) < (number[k], k)$$

En effet, pour que le processus  $P_i$  se trouve en SC il doit avoir lu  $number[k]=0$  ou  $(number[i], i) < (number[k], k)$

# Preuve de l'algorithme

**Cas 1:**  $P_i$  lit  $\text{number}[k]=0$ . Alors, lors de la lecture  $P_k$  n'a pas encore choisi un numéro.

- **Cas 1.1:**  $P_k$  n'exécute pas la phase 1. du protocole d'entrée, alors  $P_k$  va lire  $\text{number}[i]$  et choisir  $\text{number}[k] > \text{number}[i]$ .
- **Cas 1.2:**  $P_k$  exécute la phase 1. du protocole d'entrée. L'entrée de  $P_k$  en phase 1. doit être postérieure au test par  $P_i$  de  $\text{choosing}[k]$ . Alors  $P_k$  va lire  $\text{number}[i]$  et choisir  $\text{number}[k] > \text{number}[i]$

# Preuve de l'algorithme

**Cas 2:**  $P_i$  lit  $(\text{number}[i], i) < (\text{number}[k], k)$

Lors de l'entrée par le processus  $i$  en SC on a bien  $(\text{number}[i], i) < (\text{number}[k], k)$ .

En SC  $P_i$  ne modifie pas la valeur de  $\text{number}[i]$ .  $P_k$  modifie  $\text{number}[k]$  pour entrer en SC, mais cette valeur peut seulement croître.

# preuve de l'algorithme

**2<sup>ème</sup> Assertion:** Si  $P_i$  est en SC alors  $\text{number}[i] > 0$ .

En effet,  $P_i$  exécute  $\text{number}[i]++$  avant d'entrer en SC.

On montre que deux processus  $P_i$  et  $P_k$  ne peuvent pas se trouver simultanément en SC. En effet, on a  $\text{number}[i] > 0$  et  $\text{number}[k] > 0$  par la deuxième assertion.

On doit donc avoir

$(\text{number}[i], i) < (\text{number}[k], k)$  et

$(\text{number}[k], k) < (\text{number}[i], i)$

mais l'ordre est total, c'est une contradiction, l'algorithme assure donc l'exclusion mutuelle.

# Preuve de l'algorithme

## Insuffisance des ressources (starvation):

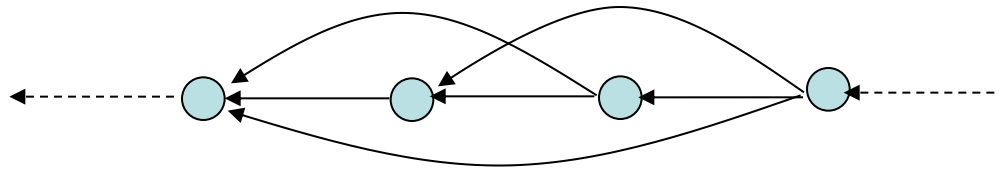
Un processus qui désire entrer en SC possède un numéro, et un nombre fini de processus se trouvent avec un numéro inférieur. Le processus aura donc accès à la SC en un temps fini.

## Remarques:

1. Les variables sont toujours modifiées par un seul processus,  $P_i$  modifie `number[i]` et `choosing[i]`
2. Le nombre d'opérations est proportionnel au nombre de processus  $O(N)$ .
3. Les numéros attribués aux processus ne sont pas bornés.

# Estampilles temporelles

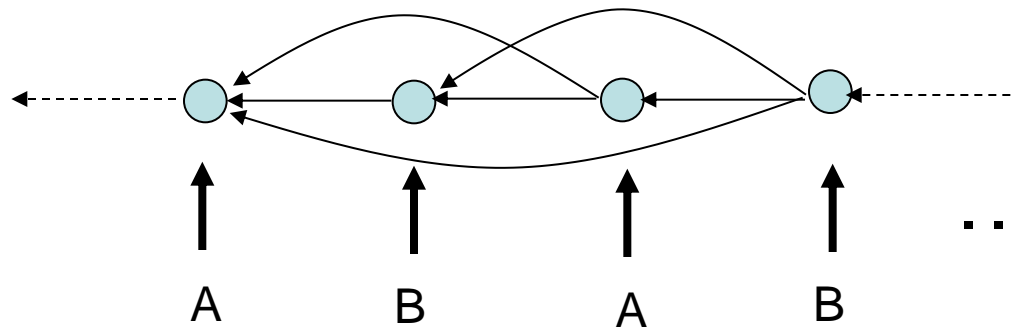
L'algorithme de Lamport utilise des compteurs *number[]* qui déterminent un ordre d'arrivée des processus. Schématiquement cet ordre se représente comme:



où chaque nœud représente une valeur de *number[]* (estampille temporelle, timestamp) et une flèche qui part d'un nœud *u* vers un nœud *v* indique que *u* est un est 'plus grand' que *v*, *u* se produit après. L'algorithme de Lamport utilise une infinité d'estampilles temporelles. Dans certaines situation un dépassement de capacité peut se produire.

# Estampilles temporelles

On considère deux processus A et B qui choisissent alternativement un nombre à chaque fois supérieur au nombre choisi par l'autre processus



On a possibilité de dépassement de capacité car le graphe de précedence utilisé est infini.

Pour résoudre ce problème on utilise un graphe de précedence fini.

# Estampilles temporelles

Pour résoudre ce problème difficile (**utiliser un nombre fini d'estampilles**), il faut d'abord modifier l'algorithme de Lamport de telle manière que la notion de précédence ne soit pas un ordre total. On peut remplacer la boucle d'attente

```
for (int j = 0; j < N; j++) {  
    while (choosing[j]); // attente proc. en phase 1.  
    while ((number[j] != 0) && ((number[j] < number[i]) ||  
        ((number[j]==number[i]) && j < i))) ; // attente active  
}  
par (pseudo-code)  
while(choosing[k] || (number[k],k)<<(number[i],i));
```



# Estampilles temporelles

de plus, on remplace le précédent code pour déterminer le numéro

```
for (int j = 0; j < N; j++)  
    if (number[i] > number[j])  
        number[i] = number[j];  
number[i]++;
```

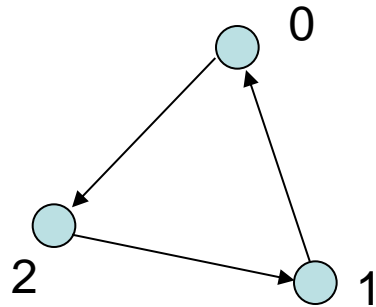
par (pseudo-code)

*number[i] = max + 1(number[0], ..., number[N - 1]);*

Dans cette version number[] n'est plus un nombre entier, c'est un nœud d'un graphe ....

# Estampilles temporelles

On considère le graphe de précedence suivant:

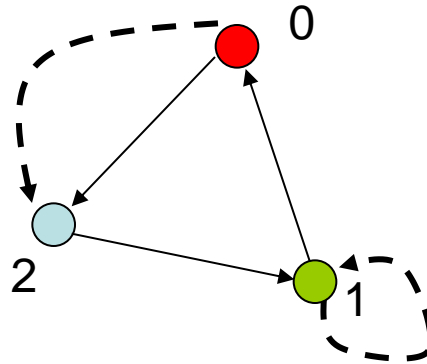


dans lequel 0 est plus petit que 1, 1 est plus petit que 2 et 2 est plus petit que 0.

**Si le nombre de processus est 2, on peut utiliser ce graphe pour générer les estampilles temporelles.**

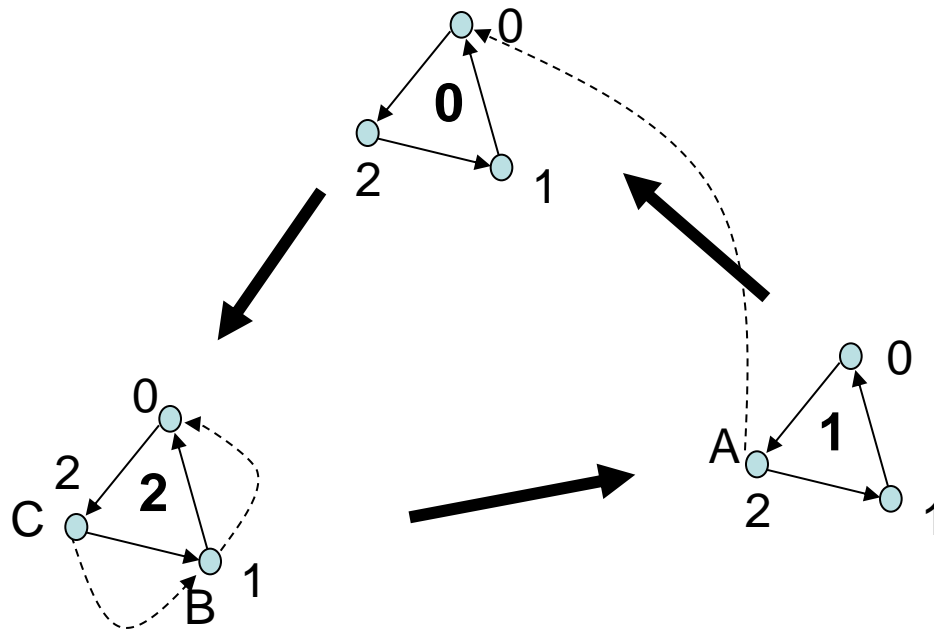
# Estampilles temporelles

En effet, supposons que le processus **A** possède l'estampille 0 et le processus **B** l'estampille 1, le numéro de **B** est plus grand.



# Estampilles temporelles

On peut généraliser le procédé pour N processus. Pour trois processus A, B, C qui se trouvent respectivement en 12, 21 et 22. B passe en 20 pour être le plus grand, puis C en 21. A passe en 00 pour être le plus grand,....



# Instructions atomiques

La difficulté principale pour assurer l'exclusion mutuelle vient du fait qu'un processus doit

1. Lire la valeur d'une variable pour tester que la SC est accessible
2. Réserver l'accès à la SC en modifiant la valeur d'une variable

et que le processus peut perdre le contrôle pendant l'intervalle de temps nécessaire à ces opérations (lecture et écriture).

En Java le mot clé **synchronized** permet de définir une routine qui ne peut être exécutée que par un seul processus simultanément.

# TestAndSet

```
public class TestAndSet {  
    int myValue = -1;
```

```
    public synchronized int testAndSet(int newValue) {  
        int oldValue = myValue;  
        myValue = newValue;  
        return oldValue;  
    }  
}
```



Section Critique,  
exclusion mutuelle,  
exécution atomique

A un instant donné, un seul processus peut exécuter la fonction testAndSet. L'environnement gère un verrou qui est attribué au processus qui exécute la fonction. Si le verrou est déjà attribué, un processus appelant est bloqué.

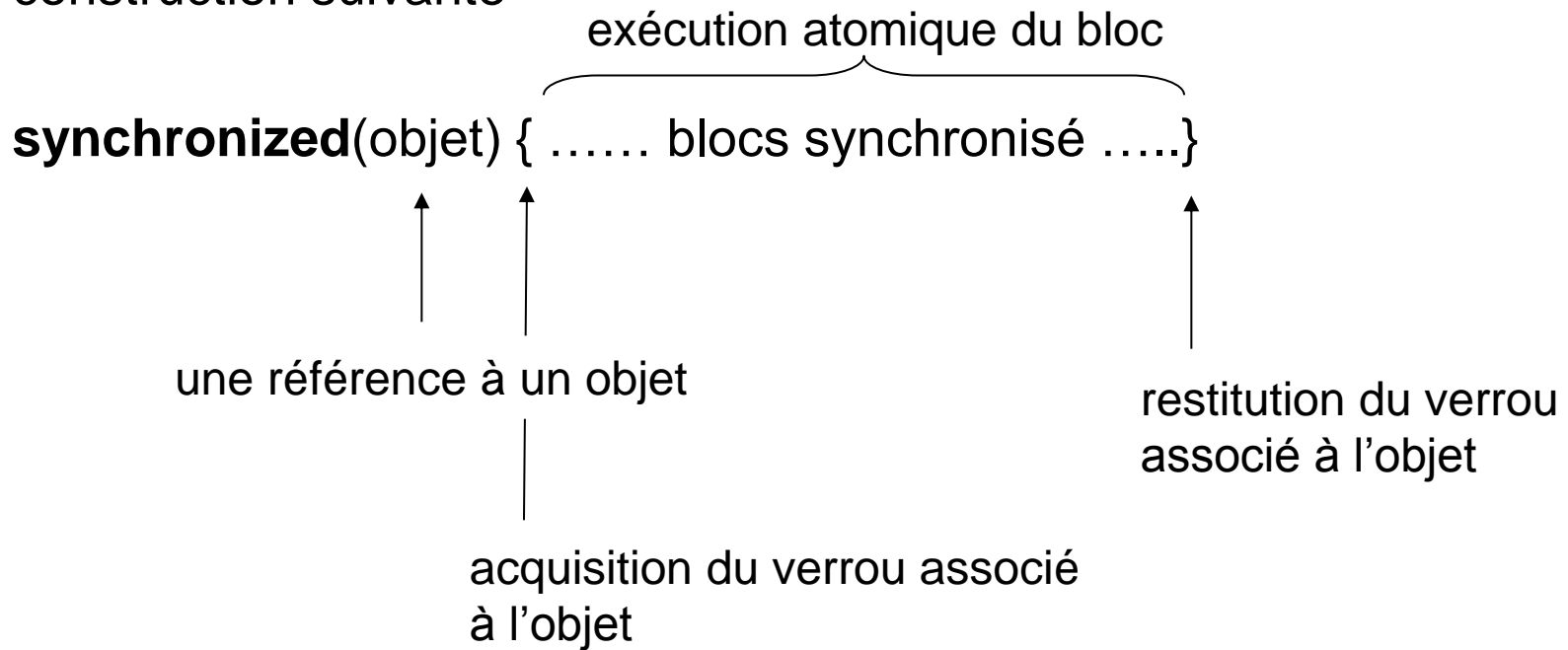
# Exclusion mutuelle avec testAndset

En utilisant testAndset on peut implémenter simplement un objet Lock

```
class HWMutex implements lock {  
    TestAndSet lockFlag;  
  
    public void requestCS(int i) { // protocole d'entrée en SC  
        while (lockFlag.testAndSet(1) == 1);  
    }  
  
    public void releaseCS(int i) { // protocole de sortie de SC  
        lockFlag.testAndSet(0);  
    }  
}
```

# Remarques

On a défini une méthode synchronisée, c'est un cas particulier de la construction suivante





# Remarques

L'utilisation de section de code synchronisée étant bloquante il faut limiter l'utilisation de ce mécanisme aux seules portions de code où c'est nécessaire. Sinon, les applications perdent en efficacité.

Java ne spécifie pas quel processus est sélectionné lorsqu'il y a contention pour un verrou. Il n'y a pas de garantie de **vivacité**.

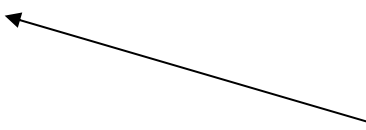
Les blocs synchronisés en java sont dits **réentrant**, c'est-à-dire que si un processus dispose du verrou sur un objet et qu'il redemande un verrou sur le même objet, il l'obtient. Ce qui veut dire que l'environnement gère l'acquisition d'un verrou et mémorise le processus qui dispose du verrou. Un processus peut donc posséder plusieurs fois un verrou, chaque opération *unlock* annule un seul *lock*.

les blocs non réentrants peuvent être la source d'inter blocages.

# Remarques

```
public class Widget {  
    public synchronized void doSomething() { ....}  
}  
}
```

```
public class LogginWidget extends Widget {  
    public synchronized void doSomething {  
        System.out.println(.....);  
        super.doSomething();  
    }  
}
```



cet appel serait la source d'un  
inter blocage si le verrou n'était  
par réentrant.

# Remarques

Un verrou est associé à chaque objet. Pour implémenter une sémaphore binaire on peut donc utiliser n'importe quel objet dérivé de la class *Object*.

```
Object obj = new Objetc();
```

Ensuite, on synchronise le code de la section critique

```
synchronized (obj) {  
    ....  
    section critique  
    ....  
}
```

# Remarques

Si une méthode synchronisée est **static** le verrou est associé à la classe. Il y a un seul verrou commun à tous les objets instances de la classe.

Lorsque l'exécution d'une méthode se termine normalement ou pas, une opération *unlock()* est automatiquement générée.

# Exclusion mutuelle en pratique

En pratique pour garantir l'exclusion mutuelle on utilise une queue FIFO.

Une solution simple pour implémenter une telle queue dans un environnement concurrent est de synchroniser les méthodes de gestion de la queue.

```
public class queue {  
    private int head = 0, tail = 0;  
    Item[QSIZE] items;  
    public synchronized void enq(Item x) {  
        while (this.tail - this.head == QSIZE) // on attend si la queue est pleine  
            this.wait();  
        this.items[this.tail++] = x; // on insère l'élément  
        this.notifyAll(); // on informe les processus bloqués après deq  
    }
```

on acquiert le verrou avant d'accéder,  
on le libère après

on incrémente après évaluation

# Queue bloquante

Utiliser des méthodes synchronisées permet d'assurer le bon fonctionnement du programme.

Il y a des situations où ce n'est pas obligatoire, par exemple si un seul processus appelle *enq()* et un seul processus appelle *deq()*.

Synchroniser les processus réduit les performances d'un algorithme.

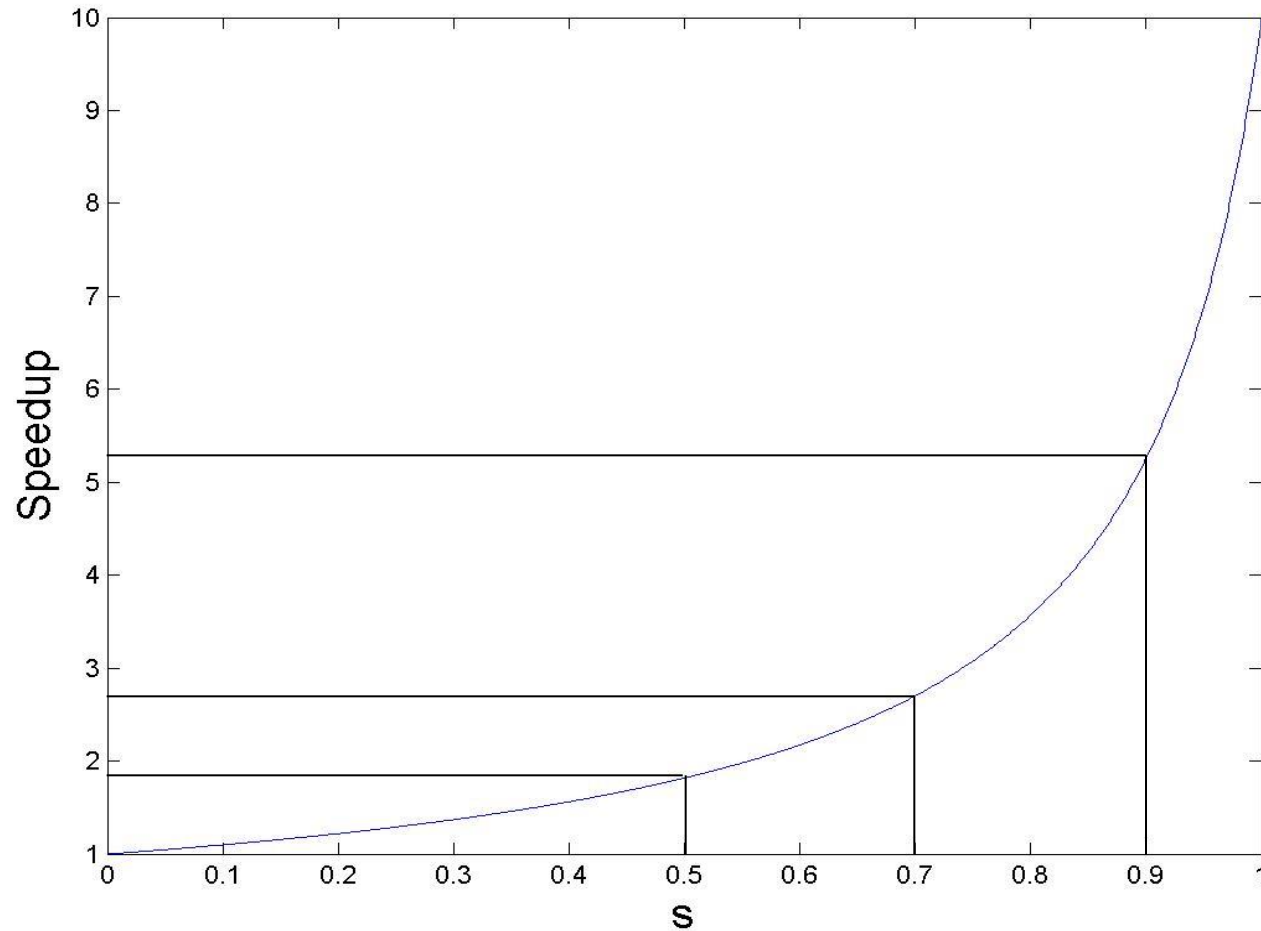
**Rappel:** La loi d'Amdahl

Soit  $s$  la proportion de programme parallélisable, et  $N$  processus à disposition

$$speedup = \frac{1}{1 - s + \frac{s}{N}}$$

# Queue bloquante

loi d'Amdhal 10 processeurs



# Queue bloquante

On observe que le speedup diminue très rapidement en fonction de la proportion de code parallèle. Les méthodes *synchronisées* ne sont pas exécutables en parallèle et pénalisent les performances du système.

Des objets complexes sont implémentés en utilisant d'autres mécanismes qui leur permettent d'être accédés par plusieurs processus simultanément sans mécanisme de blocage (voir plus tard, lock-free algorithm).

Dans un système réel les processus sont susceptibles d'être interrompus, par exemple pour le traitement d'une exception. C'est particulièrement gênant si le processus interrompu exécute une méthode synchronisée.

Néanmoins, il est important d'utiliser les mécanismes de synchronisations pour s'assurer du bon fonctionnement du programme.



# Preuves des algorithmes concurrents

Pour les algorithmes présentés on a développé des preuves qui sont basées sur l'analyse des différentes exécutions possibles (cas1, cas2, ...). Ce type d'analyse peut être formalisé en utilisant les diagrammes d'états et de transitions. On a deux difficultés avec cette approche:

1. Il est difficile de se persuader que l'on a pas 'oubliée' de traiter une situation, la construction du diagramme est problématique dès que le nombre d'état devient grand. Les exemples ont principalement montrés des fragments de diagrammes qui montrent que les algorithmes ne fonctionnent pas (cas particulier). Cette approche est problématique pour montrer des propriétés qui ne doivent jamais se produire (safety).

# Preuves des algorithmes concurrents

2. La méthode est difficilement 'mécanisable', par exemple si on veut développer un programme qui vérifie automatiquement qu'une propriété donnée est toujours vérifiées.

De plus, trouver les erreurs dans les programmes concurrents par essai-erreur-correction (test) est souvent difficile voir impossible:

1. Les erreurs peuvent se produire seulement dans des conditions très particulières (liées à la course entre les données, data race) qui peuvent se produire tous les mois, les années,...
2. Quelque fois les mécanismes d'observations ne permettent plus aux exécutions fautives d'exister...
3. Comment peut-on décider que plusieurs processus sont inter bloqués? (il faut définir une période de temps)

# Preuve des algorithmes concurrents

Pour toutes ces raisons, et d'autres, des méthodes formelles ont été développées et sont essentielles pour la validation des programmes concurrents. Des programmes tels que SPIN sont des outils d'aide à la validation.

Ces méthodes font l'objet d'autres cours (par exemple D.Buchs, concurrency and distribution).

Une méthode intermédiaire pour démontrer formellement des propriétés des programmes concurrents est d'utiliser des invariants.

# Invariants

On utilise exclusivement la logique des propositions.

On a un ensemble de propositions atomiques  $\{p, q, r, \dots\}$

Les opérateurs,  $\neg$  (non),  $\vee$  (ou),  $\wedge$  (et),  $\Rightarrow$  (implication),  $\Leftrightarrow$  equivalence.

Les propositions atomiques sont des expressions constituées des variables et des pointeurs de programmes.

# Invariants

Par exemple, pour la deuxième tentative

```
boolean wantCS[] = {false, false};
```

1 **section non critique**

```
    public void requestCS(int i) {
```

2 wantCS[i] = **true**;

3 **while** (wantCS[1-i]); }

4 **section critique**

```
    public void releaseCS(int i) {
```

5 wantCS[i] = **false**;

wantCS[0], wantCS[1], sont des propositions atomiques

wantCS[1] ^ p4 (le processus p exécute la ligne 4 et wantCS[1]=true, p  
le processus numéro 0 et q le processus numéro 1)

Pour la troisième tentative, on a turn=1 est une proposition atomique.

# Invariant - définition

## Définition:

Une proposition est invariante si elle est toujours vraie, c'est-à-dire dans tous les états possibles du programme.

Pour prouver que la deuxième tentative pour résoudre l'exclusion mutuelle est correcte, on doit prouver que:

$$\neg(p4 \wedge q4)$$

est toujours vraie (c'est un invariant). Une telle preuve se fait par induction

1. On prouve qu'elle est vraie initialement
2. On suppose l'assertion vraie dans tous les états jusqu'à l'état courant et on montre qu'elle est vraie dans tous les états suivants possibles.

# Preuve de l'exclusion mutuelle avec les invariants

**Lemme 1:**  $p3 \vee p4 \vee p5 \Rightarrow \text{wantCS}[0]$  est un invariant

- a. Dans l'état initial on a  $\text{wantCS}[0]=\text{false}$ , alors la proposition est vraie.
- b. (induction) Aucune exécution du processus  $q$  ne peut modifier la validité de la proposition car elle dépend que de l'état courant de  $p$  et de  $\text{wantCS}[0]$  qui est modifiée que par  $p$ .

La proposition ne peut être fausse que lorsque  $p$  exécute  $p3$ ,  $p4$  ou  $p5$  ( $1 \Rightarrow 0$ ).

Exécuter  $p3$  ou  $p4$  ne modifie pas la validité de l'assertion car dans les deux cas  $p3 \vee p4 \vee p5$  et  $\text{wantCS}[0]$  ne sont pas modifiés

Lorsque  $p5$  est exécutée alors  $p3 \vee p4 \vee p5 = \text{false}$  et l'assertion est toujours valide ( $0 \Rightarrow x$  est vraie).

Lorsque  $p2$  est exécutée alors  $p3 \vee p4 \vee p5 = \text{true}$  (le compteur de programme devient  $p3$ , mais on a bien  $\text{wantCS}[0]=\text{true}$ ).

# Preuve de l'exclusion mutuelle avec les invariants

En fait, seuls p2 et p5 peuvent modifier la validité de la proposition.

**Lemme 2:**  $\text{wantCS}[0] \Rightarrow p3 \vee p4 \vee p5$  est un invariant.

En effet, cette proposition peut être fausse que si  $\text{wantCS}[0]=\text{true}$ .

- a. Dans l'état initial  $\text{wantCS}[0]=\text{false}$ , la proposition est donc vraie
- b. Les seules exécutions qui peuvent modifier la validité de la proposition sont p2 et p5. Après p2 on a  $\text{wantCS}[0] = \text{true}$  mais le compteur de programme se trouve en p3, la proposition est valide. Après p5  $\text{wantCS}[0]=\text{false}$ ,...



# Preuve de l'exclusion mutuelle avec les invariants

**Lemme 3:**  $p3 \vee p4 \vee p5 \Leftrightarrow \text{wantCS}[0]$  et  $q3 \vee q4 \vee q5 \Leftrightarrow \text{wantCS}[1]$

Les deux lemmes précédents montrent l'équivalence et il est clair que les mêmes preuves s'appliquent au processus q.

**Théorème:**  $\neg(p4 \wedge q4)$  est invariant.

a. L'assertion est trivialement vraie initialement.

b. Seules deux exécutions doivent être vérifiées. Si p3 est exécutée avec succès lorsque q4 est vraie, et q3 est exécutée avec succès lorsque p4 est vraie (symétrique).

pour que p3 soit exécutée avec succès, il faut que  $\text{wantCS}[1] = \text{false}$ .  
D'après le lemme précédent q4 ne peut pas être vraie, c'est une contradiction et p3 ne peut pas s'exécuter avec succès.

# Remarques

Les logiques temporelles (LTL, CTL,...) permettent de généraliser l'approche basée sur les invariants.

Il est difficile de prouver des conditions de vivacité (liveness), c'est-à-dire qu'un événement souhaité se produira nécessairement avec les invariants. En logique temporelle on a un opérateur dédié.

On a toujours supposé que considérer toutes les différentes exécutions possibles revient à entrelacer (interleaving) les différentes instructions à exécuter, c'est-à-dire

p1 p2 q1 p3 p4 q2 ,...

p1 p2 p3 p4 p5 q1 q2, ...

etc.

# Synchronisation

# Mécanismes de synchronisations

Les mécanismes présentés dans la première partie réalisent l'exclusion mutuelle et utilisent des boucles d'attente actives.

Des mécanismes de synchronisations permettent

1. de réaliser l'exclusion mutuelle
2. de coordonner l'exécution des threads

sans utiliser des boucles d'attentes actives.

# Sémaphore booléenne

Une sémaphore booléenne est un objet qui permet de réaliser l'exclusion mutuelle et qui dispose de deux champs de données

1. une valeur booléenne
2. une file d'attente de processus bloqués

et deux méthodes P() et V().

Lorsque P() est exécutée on a

1. si valeur=true, alors valeur=false
2. si valeur=false, le processus appelant est bloqué

Lorsque V() est exécutée on a

1. valeur=true, si la file d'attente des processus n'est pas libre on libère un processus

# Sémaphore booléenne

```
public class Binarysemaphore {  
    private boolean value;  
    BinarySemaphore(boolean initValue) {  
        value = initValue;  
    }  
    public synchronized P() { // protocole d'entrée en SC  
        while (value == false)  
            try { this.wait() } catch (InterruptedException e) {}  
        value = false;  
    }  
    public synchronized V() { // protocole de sortie de SC  
        value = true;  
        notify();                // libère un processus en attente  
    }  
}
```

# SC avec sémaphore

Pour assurer l'exclusion mutuelle on utilise le code suivant

```
BinarySemaphore mutex = new BinarySemaphore(true);
```

```
mutex.P();
```

```
section critique
```

```
mutex.V();
```

# Remarques

Les méthodes `wait()` et `notify()` sont des méthodes de la classe `Object`.

Ces méthodes peuvent être exécutées seulement depuis des méthodes qui disposent du verrou sur l'objet, c'est-à-dire qui sont **synchronisées**. Dans le cas contraire une exception *`IllegalMonitorStateException`* est levée.

La méthode `wait()` bloque le processus appelant et libère le verrou sur l'objet. Attention, si le processus dispose de plusieurs verrous, seul le verrou associé à l'objet auquel appartient la méthode depuis laquelle `wait()` est exécutée est libéré. **Possibilité d'inter blocage.**



# Remarques

La méthode `notify()` active un processus en attente, le langage Java ne spécifie pas lequel.

Il existe une spécification temps-réel de java (RTSJ) qui spécifie l'ordre dans lequel les processus doivent être libérés.

Lorsque `notify()` est exécutée, le verrou sur l'objet n'est pas libéré (le thread qui s'exécute continue de s'exécuter).

La méthode `notifyAll()` permet d'activer tous les processus en attente, ne libère pas le verrou et les processus activés doivent essayer d'acquiescer le verrou avant de continuer.

`notify()` et `notifyAll()` n'ont pas d'effet si aucun processus n'est en attente.

# Remarques

Un processus peut être réactivé en utilisant `Thread.interrupt()`

Lorsqu'un processus est réactivé, il devient exécutable mais ne s'exécute pas nécessairement immédiatement. C'est-à-dire que la condition qui a permis sa réactivation n'est pas forcément vérifiée lorsqu'il s'exécute. Dans notre implémentation des sémaphores:

**while** (value == **false**)

**try** { this.wait } **catch** (InterruptedException e) {}

Lorsque le processus est réactivé `value==false` est possible.

Pour cette raison, le code suivant **ne fonctionne pas**

**if** (value == **false**)

**try** { this.wait } **catch** (InterruptedException e) {}

# Remarques

les méthodes `Thread.sleep()` et `Thread.yield()` permettent à un thread de forcer l'ordonnanceur à changer le thread actif. Néanmoins,

1. un thread ne perd pas les verrous dont il dispose.
2. ces opérations n'ont pas d'effets sur la synchronisation; les données sauvegardées dans les registres ne sont pas copiées en mémoire principale; les données ne sont pas lues depuis la mémoire principale après l'exécution.

# Retour sur *wait()*

Il y a trois versions:

*wait()*, *wait(long millisecs)*, *wait(long millisecs, int nanosecs)*.

Si la valeur des paramètres est nulle, les appels sont équivalents à *wait()*.

La méthode *wait()* peut lever une exception `InterruptedException`

Soit un thread **t** qui exécute une méthode *wait()* dans l'objet **m** et **n** est le nombre de *lock()* de **t** sur **m**.

- si **n** = 0 (pas de *lock()* ) une exception **`IllegalMonitorStateException`** est levée
- si l'argument nanosecs n'est pas dans l'intervalle 0-999999, ou l'argument millisecs est négatif l'exception **`IllegalArgumentException`** est levée

# Retour sur *wait()*

- Si le thread **t** est **Interrupted** une exception **InterruptedException** est levée

Sinon la séquence suivante se produit

- Le thread **t** est ajouté à l'ensemble des threads en attente de l'objet **m** et exécute **n** opérations *unlock()* (libère le verrou) sur l'objet **m**.
- Le thread est bloqué jusqu'à être retiré de l'ensemble des processus en attente dans **m**. Ce qui se produit si:
  - *notify()* est exécuté dans **m** et **t** est sélectionné.
  - *notifyAll()* est exécuté dans **m**.
  - *t.interrupt()* est exécutée.
  - Le temps d'attente spécifié est écoulé.

# Retour sur *wait()*

- Le thread **t** exécute **n** *lock()* sur **m** (une fois activé).
- Si **t** à été retiré de l'ensemble des processus en attente après exécution de *t.interrupt()*, l'exception **InterruptedException** est levée.

# Retour sur *wait()*

La spécification Java laisse aux implémentations de JVM la possibilité de générer une action interne pour retirer un processus en attente sans qu'aucune des conditions citées précédemment ne se produisent. Cette pratique n'est pas recommandée, néanmoins il est spécifié de placer le *wait()* dans une boucle *while* et de tester une condition logique si nécessaire.

# Retour sur *notify()*

De même, pour *notify()*, on a

- si  $n = 0$  l'exception **IllegalMonitorStateException** est levée (le thread  $t$  ne possède pas le verrou sur  $m$ )
- si  $n > 0$  et que l'ensemble des processus en attente n'est pas vide, un processus est sélectionné et devient exécutable (*notify()*) ou idem pour tous les threads si *notifyAll()*.



# Retour sur interrupt()

Lorsqu'un thread **t** exécute **u.interrupt()** (**u** et **t** pas forcément différents) le drapeau d'état du processus **u** est positionné.

Si **u** est en attente dans un objet **m** il est retiré de l'ensemble des processus en attente. Après avoir obtenu le verrou sur l'objet, il lève l'exception **InterruptedException**.

# Remarques

Lorsqu'un thread a exécuté avec succès *semaphore.P()*, il est important qu'il exécute *semaphore.V()* en quittant la section critique pour permettre à un autre thread d'y accéder. Pour s'assurer que le thread ne bloque pas d'autres threads **si une exception est levée**, on accède la section critique comme cela:

```
semaphore.P();
```

```
try {
```

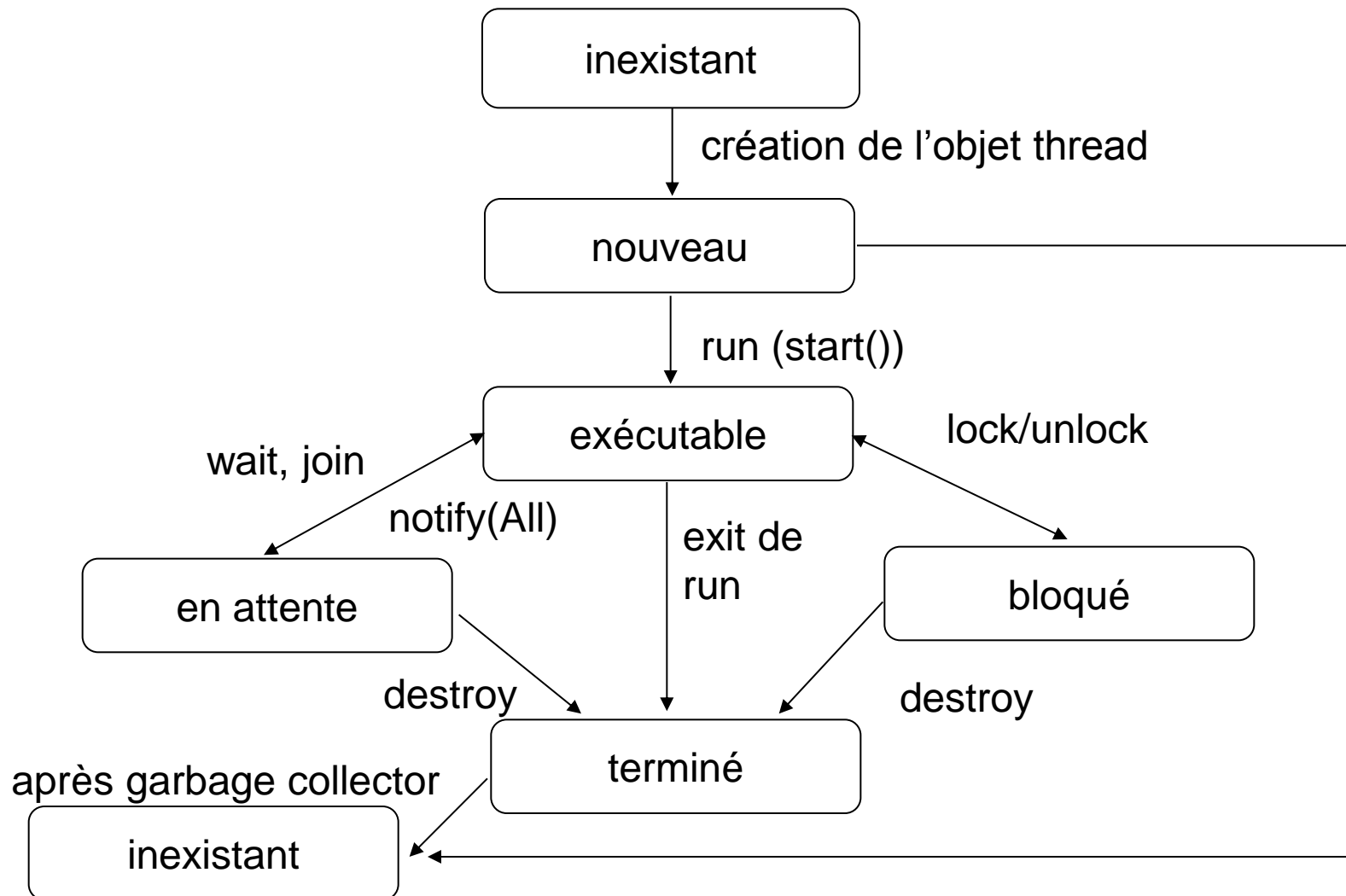
```
    ... section critique...
```

```
} finally {
```

```
    semaphore.V(); // cette instruction est toujours exécutée, qu'une
```

```
}                // exception ait été levée ou non
```

# Les différents états d'un thread



# Remarques

Les méthodes wait(), notify(), notifyAll() permettent de synchroniser les processus.

Il est aussi possible pour un thread d'attendre sur la fin (thread dans l'état terminé) d'un autre thread en utilisant la méthode join().

```
public class Exemple extends Thread { ....}
```

```
public static void main(....) {  
    Exemple objet = new Exemple(....);  
    objet.start()  
    try {  
        objet.join();  
    } catch (InterruptedException e){};  
}
```

# Sémaphore entière

```
public class CountingSemaphore {  
    private int value;  
    public CountingSemaphore(int initValue) {  
        value = initValue;  
    }  
    public synchronized void P() {  
        while (value == 0) {  
            try { this.wait(); } catch (InterruptedException e) {}  
            value--;  
        }  
    }  
    public synchronized void V() {  
        value++;  
        if (value == 1) notify();  
    }  
}
```

Le thread libéré ne s'exécute pas nécessairement



# Invariants pour les sémaphores

On considère seulement les sémaphores entières, une sémaphore booléenne étant un cas particulier de sémaphore entière.

- Une sémaphore  $s$  est initialisée avec la valeur `initValue`.
- La valeur de la sémaphore  $s.value$  satisfait toujours

$$s.value \geq 0$$

- Si on note  $\text{comp\_P}(s)$  et  $\text{comp\_V}(s)$  le nombre d'exécution complète des méthodes  $P()$  et  $V()$  de la sémaphore  $s$ , on a toujours

$$s.value = \text{initValue} - \text{comp\_P}(s) + \text{comp\_V}(s)$$

# Preuve de SC

On suppose qu'une section critique est accédée par tous les processus selon le protocole déjà vu:

```
s.P();  
    section critique  
s.V();
```

et que s est initialisée avec  $\text{initValue} = 1$ .

Supposons que deux processus se trouvent en section critique simultanément. Comme  $s.P() \longrightarrow \text{section critique} \longrightarrow s.V()$ , on a  $-\text{comp\_P}() + \text{comp\_V}() \leq -2$ .

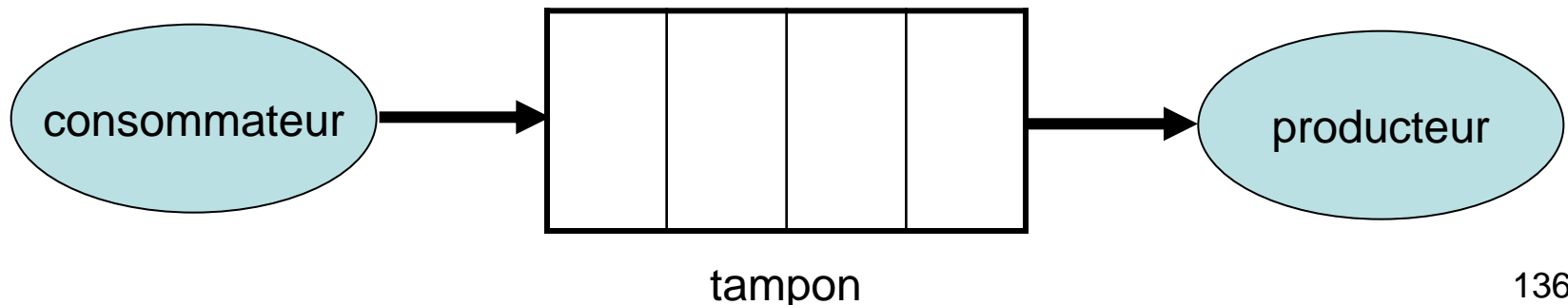
On a alors une contradiction avec les invariants associés à la sémaphore car  $\text{cars.value} = \text{initValue} - \text{comp\_P}() + \text{comp\_V}() \leq -1$

# Le problème du producteur-consommateur

Le problème du producteur-consommateur modélise des problèmes concrets dans lesquels on peut identifier deux types de comportements

1. Un processus (producteur) qui détermine les tâches à exécuter
2. Un processus (consommateur) qui exécute les tâches.

Les deux processus communiquent en déposant/retirant des objets d'un tampon de taille finie.





# Le problème du producteur-consommateur

une solution de ce problème consiste à définir un thread producteur et un thread consommateur.

La programmation concurrente s'impose pour résoudre ce problème car:

1. Le problème s'exprime naturellement avec des threads.
2. Les programmes relatifs au producteur/consommateur peuvent être indépendants. Le producteur n'a pas besoin de connaître comment est implémenté le consommateur.
3. Avec la programmation concurrente on ne doit pas considérer explicitement les problèmes de vitesses différentes des processus.

# Le problème du producteur-consommateur

On trouve ce genre de problèmes, par exemple:

1. Accès à une ressource séquentielle, telle qu'une imprimante ou des périphériques d'entrée/sortie.
2. Un programme qui scan un disque à la recherche de fichiers et les indexe pour les retrouver plus rapidement ensuite. Le producteur scan le disque, le consommateur les indexe.
3. Un browser web produit des informations qui doivent être transmises par le consommateur via une ligne de communication
4. Le clavier produit des données qui doivent être transmises à la tâche concernée.
5. Un jeu produit des images qui doivent être affichées (consommateur)
6. ....

# Le problème du producteur-consommateur

Les contraintes de synchronisation pour ce problème sont:

- Le producteur ne doit pas ajouter un élément si le buffer est plein.
- Le consommateur ne doit pas retirer un élément si le tampon est vide.

On parle de **synchronisation conditionnelle**, un processus doit attendre qu'une condition soit satisfaite pour continuer de s'exécuter.

# Code java – le tampon avec sémaphores

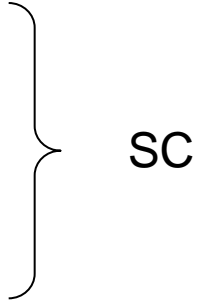
```
class BoundedBuffer {  
    final int size = 10; // taille du tampon  
    double[] buffer = new double[size]; // tampon  
    int inBuf =0, outBuf = 0; // pointeurs  
    BinarySemaphore mutex = new BinarySemaphore(true);  
    CountingSemaphore isEmpty = new CountingSemaphore(0);  
    CountingSemaphore isFull = new CountingSemaphore(size);  
    ...
```

pour l'accès en  
section critique

pour la synchronisation conditionnelle

# Code Java – le tampon avec sémaphores

```
public void deposit (double value) {  
    isFull.P();           //attente si le tampon est plein  
    mutex.P();            // accès exclusif au tampon  
    buffer[inBuf] = value; // on place la valeur  
    inBuf = (inBuf + 1) % size;  
    mutex.V();            // on quitte la section critique  
    isEmpty.V();          // notification au consommateur  
}
```



...

# Code Java – le tampon avec sémaphores

```
public double fetch() {  
    double value;  
    isEmpty.P();    // attente si buffer vide  
    mutex.P();      // accès exclusif au tampon  
    value = buffer[outBuf];  
    outBuf = (outBuf + 1) % size;  
    mutex.V();      // on quitte la section critique  
    isFull.V();     // notification au producteur  
    return value;  
}  
}
```

# Utilisation du tampon fini

```
import java.util.Random;
class Producer implements Runnable {
    BoundedBuffer b = null;
    public Producer(BoundedBuffer initb) {
        b = initb;
    }
    public void run() {
        double item;
        Random r = new Random();
        while(true) {
            item = r.nextDouble();
            System.out.println("valeur produite " + item);
            b.deposit(item);
            sleep(50); // cette instruction ne fonctionne pas comme cela....
        }
    }
}
```

# Utilisation du tampon fini

```
class Consumer implements Runnable {  
    BoundedBuffer b = null;  
    public Consumer(BoundedBuffer initb) {  
        b = initb;  
    }  
    public void run() {  
        double item;  
        while (true) {  
            item = b.fetch();  
            System.out.println("valeur à consommer " + item);  
            sleep(50);    // cette instruction ne fonctionne pas comme cela...  
        }  
    }  
}
```



# Utilisation du tampon fini

```
class ProducerConsumer {  
    public static void main(String [] args) {  
        BoundedBuffer buffer = new BoundedBuffer();  
        Producer producer = new Producer(buffer);  
        Consumer consumer = new Consumer(buffer);  
        new Thread(producer).start();  
        new Thread(consumer).start();  
    }  
}
```

# Preuve de la synchronisation

On considère les méthodes *deposit(...)* et *fetch()*.

On a déjà montré que l'utilisation de la sémaphore *mutex* assure l'exclusion mutuelle lors des accès au tampon.

*deposit(..)*

...

mutex.P();

buffer[inBuf] = value;

inBuf = (inBuf + 1) % size;

mutex.V();

*fetch()*

...

mutex.P();

value = buffer[outBuf];

outBuf = (outBuf + 1) % size;

mutex.V();

# Preuve de la synchronisation

On montre que la sémaphore *isEmpty* assure que l'on a jamais **underflow**, c'est-à-dire que le consommateur n'accède pas à plus de données que le producteur a déposé.

On sait par invariance que

$$\text{isEmpty.value} = \text{initvalue} - \text{comp\_P}(\text{isEmpty}) + \text{comp\_V}(\text{isEmpty}) \geq 0$$

ou encore

$$\text{comp\_V}(\text{isEmpty}) \geq \text{comp\_P}(\text{isEmpty})$$

La méthode *deposit(..)* se termine par *isEmpty.V()* et la méthode *fetch()* commence par *isEmpty.P()*. L'inégalité ci-dessus nous indique que le nombre d'exécutions de *deposit()* (terminée) est toujours plus grand ou égal au nombre d'exécutions de *fetch()* terminée ou en cours. Il ne peut donc pas y avoir d'underflow.

# Preuve de la synchronisation

On montre que la sémaphore *isFull* assure que l'on a jamais **overflow**, c'est-à-dire que le producteur ne dépose pas plus de données que peut contenir le tampon (size).

On sait par invariance que

$$\text{isFull.value} = \text{size} - \text{comp\_P}(\text{isFull}) + \text{comp\_V}(\text{isFull}) \geq 0,$$

ou encore

$$\text{size} \geq \text{comp\_P}(\text{isFull}) - \text{comp\_V}(\text{isFull}) .$$

La méthode *deposit()* commence par *isFull.P()* et la méthode *fetch()* se termine par *isFull.V()*.

Donc le nombre d'exécutions de *deposit* en cours ou terminées moins le nombre d'exécutions de *fetch()* est toujours inférieur à la taille du tampon, il ne peut donc pas y avoir d'overflow.

# Remarque sur la synchronisation

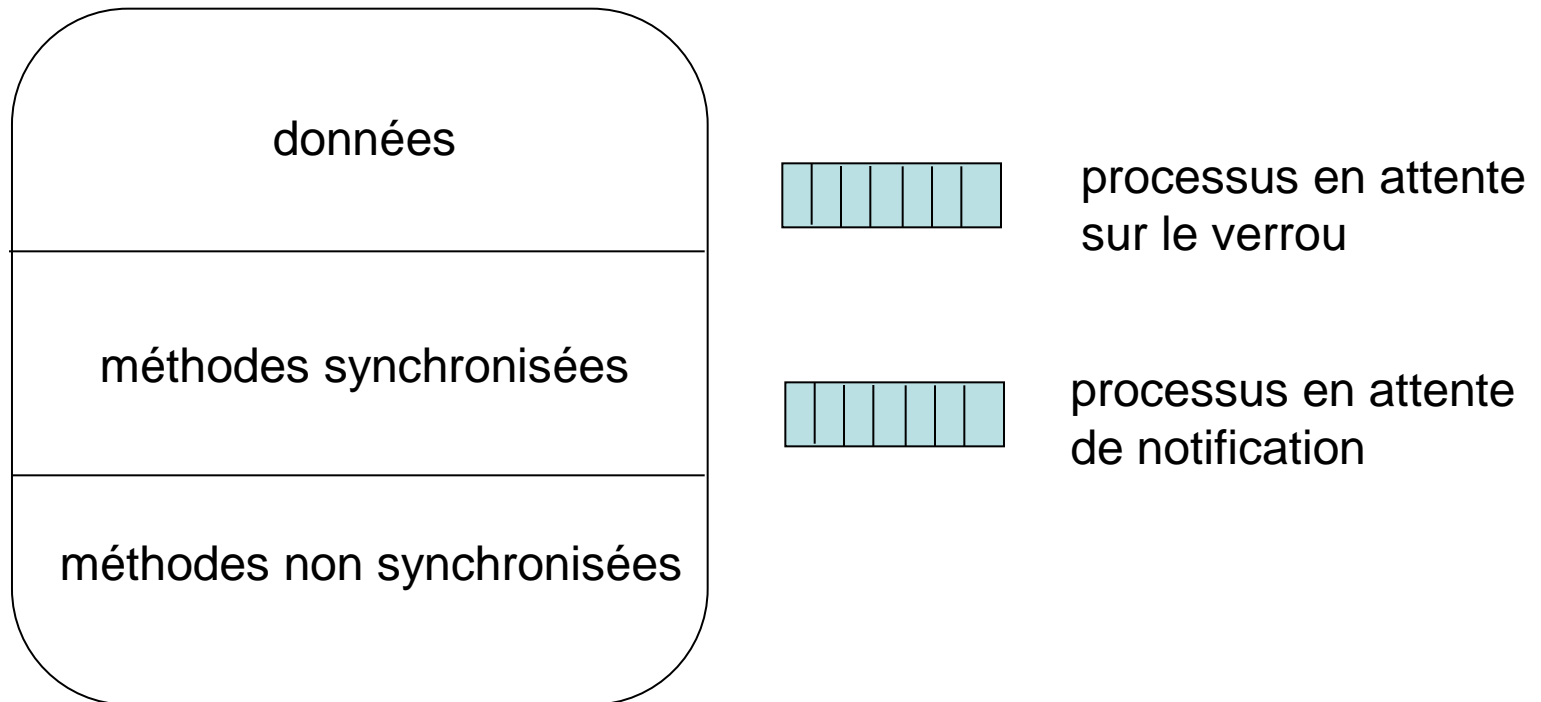
L'implémentation proposée pour le tampon de taille finie utilise une section critique pour accéder le tampon (le tableau buffer).

Si on suppose qu'il n'y a qu'un seul producteur et un seul consommateur cette section critique est superflue.

En effet, la section critique est utile seulement lorsqu'on accède une variable commune. Dans notre cas, on accède une même position du tableau seulement si  $\text{inBuf} = \text{outBuf}$ , ce qui correspond aux situations où le tableau est vide ou plein et des accès simultanés par le producteur et le consommateur sont donc impossibles.

Cette propriété se démontre formellement en utilisant les invariants des sémaphores.

# Représentation d'un moniteur Java



# Implémentation d'un tampon fini avec les moniteurs Java

```
Class BoundedBufferMonitor {
```

```
    final int sizeBuf = 10;
```

```
    double[] buffer = new double[sizeBuf];
```

```
    int inBuf = 0, outBuf = 0, count = 0;
```

```
    public synchronized void deposit(double value) {
```

```
        while (count == sizeBuf) // le tampon est plein
```

```
            myWait(this); // ne fonctionne pas comme cela
```

```
        buffer[inBuf] = value;
```

```
        inBuf = (inBuf + 1) % sizeBuf;
```

```
        count++;
```

```
        if (count == 1) notify(); // libère un éventuel thread en attente
```

```
    }
```

assure l'exclusion mutuelle  
correspond à acquérir le verrou

le thread est placé dans la file  
d'attente

# Implémentation d'un tampon fini avec les moniteurs Java

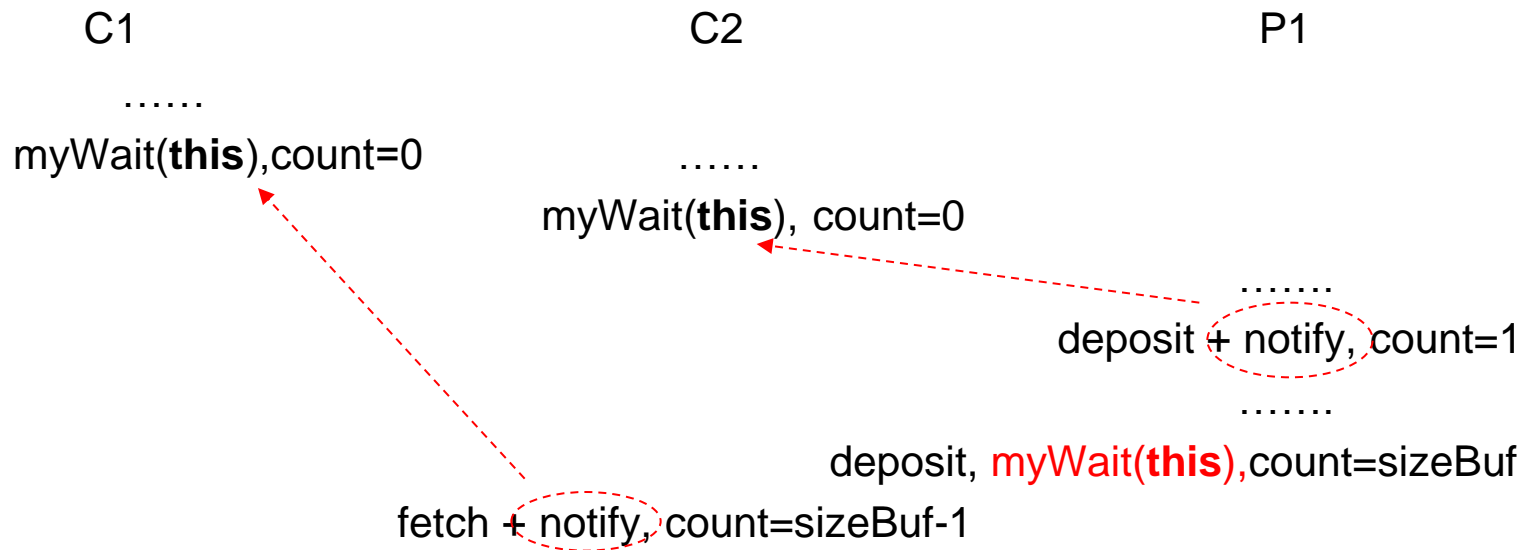
```
public synchronized double fetch();  
    double value;  
    while (count == 0) // le tampon est vide  
        myWait(this);  
    value = buffer[outBuf];  
    outBuf = (outBuf + 1) % sizeBuf;  
    count--;  
    if (count == sizeBuf - 1)  
        notify(); // libère un éventuel thread en attente  
    return value;  
}  
}
```



# Une exécution particulière

L'implémentation proposée du tampon fini est correcte seulement dans le cas d'un consommateur et un producteur.

En effet, supposons que l'on a deux consommateurs C1 et C2 et un producteur P1 et considérons l'exécution suivante.



seuls C1 et C2 sont exécutables, ils vident le tampon, exécutent myWait() et tous les processus sont en attente.

[illegible]

# Solutions

Le problème survient parce que la notification n'est pas sélective. Un processus consommateur est notifié alors que le *notify()* s'adressait au producteur.

Une première solution à ce problème est d'utiliser *notifyAll()* à la place de *notify()*.

Cette solution a le désavantage d'entraîner beaucoup de changement de contextes, certains inutiles.

Une solution plus élégante consiste à implémenter des **variables de conditions** qui permettent à un processus d'attendre qu'une condition particulière soit réalisée (par exemple, `bufferNoFull`, `bufferNotEmpty`).

# Les variables de conditions

Si on considère l'implémentation d'un tampon dans le problème du producteur-consommateur, lorsque le producteur est bloqué, il attend sur la condition `isFull==false` (`bufferNotFull`) alors que le consommateur attend sur la condition `isEmpty==false` (`bufferNotEmpty`).

Dans l'implémentation du tampon en Java, les appels à *notify()* ne sont pas sélectifs et tous les processus deviennent exécutables, qu'ils soient producteurs ou consommateurs.

# multi-producteur multi-consommateur

On définit deux objets pour la synchronisation

```
private Object conveyD = null, conveyF = null;
```

Le tampon est géré comme dans l'exemple précédent avec inBuf et outBuf et un compteur.

On utilise deux compteurs supplémentaires spaces et elements qui permettent de compter le nombre de producteur et consommateur en attente respectivement.

Initialisation: spaces = sizeBuf, elements = 0

# multi-producteur

```
public void deposit(double value) {  
    synchronized(conveyD) {  
        spaces--;  
        if (spaces < 0){ // si négatif, compte le nombre de producteur en attente  
            try {  
                conveyD.wait(); // tampon plein, attente  
            } catch (InterruptedException e) {}  
        }  
        buffer[inBuf]=value; ....  
        synchronized(conveyF) {  
            elements++;  
            if (elements <=0) conveyF.notify(); } // notify consommateur  
    }  
}
```

# multi-consommateur

```
public double fetch() {  
    double value;  
    synchronized(conveyF) {  
        elements--; // si négatif compte le nombre de consommateur en attente  
        if (elements < 0) {  
            try {  
                conveyF.wait();  
            } catch (InterruptedException e) {}  
        }  
        value = buffer[outBuf]; ....  
        synchronized(conveyD) {  
            spaces++;  
            if (spaces <= 0) conveyD.notify(); // notifie un producteur  
        }  
    }  
}
```

# Le problème des lecteurs-rédacteurs

Ce problème classique est une extension du problème de la section critique. En effet, plusieurs processus désirent accéder une même ressource mais les processus sont divisés en deux classes:

1. Les lecteurs, qui peuvent accéder la ressource de manière concurrente.
2. Les rédacteurs, qui doivent accéder la ressources comme une ressource critique.

En d'autres termes,

les lecteurs doivent exclure l'accès à la ressource aux seuls rédacteurs.

Les rédacteurs doivent exclure l'accès à la ressource aux lecteurs et aux rédacteurs.



# Le problème des lecteurs-rédacteurs

Ce problème est une abstraction des accès à une base de données partagée.

Les utilisateurs qui ne modifient pas la base peuvent l'accéder de manière concurrente.

Les modifications à la base de donnée doivent se faire de manière atomique pour éviter des problèmes de cohérence.

Le protocole de base pour accéder la ressource en lecture est d'appeler les routines *startRead()* et *endRead()*.

De même pour accéder la ressource en écriture il faut appeler les routines *startWrite()* et *endWrite()*.

# Le problème des lecteurs-rédacteurs

Plusieurs lecteurs peuvent accéder la ressource, il faut compter le nombre de lecteurs *numReaders*.

1. Pour savoir quand un rédacteur peut accéder la ressource.
2. Le premier lecteur doit empêcher l'accès à un rédacteur.
3. Le dernier lecteur à quitter la ressource doit le signaler à d'éventuels rédacteurs en attente.

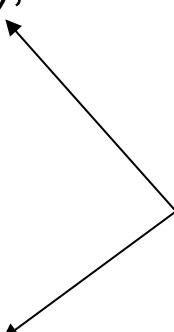
On utilise une variable entière *numReaders* pour compter les lecteurs.

Une sémaphore binaire pour restreindre l'accès à la ressource.

Une sémaphore binaire pour manipuler les données critiques du protocole. C'est-à-dire pour manipuler *numReaders*, cette sémaphore va aussi bloquer les lecteurs lorsque la ressource est accédée par un rédacteur.

# Le problème des lecteurs-rédacteurs

```
class ReaderWriter {  
    int numReaders = 0;  
    BinarySemaphore mutex = new BinarySemaphore(true);  
    BinarySemaphore wlock = new BinarySemaphore(true);  
    public void startRead() {  
        mutex.P(); // accès aux variables critiques du protocole  
        numReaders++;  
        if (numReaders == 1) wlock.P(); // réserve l'accès à la ressource  
        mutex.V();  
    }  
    public void endRead() {  
        mutex.P();  
        numReaders--;  
        if (numReaders == 0) wlock.V(); // libère l'accès à la ressource  
        mutex.V();  
    }  
}
```



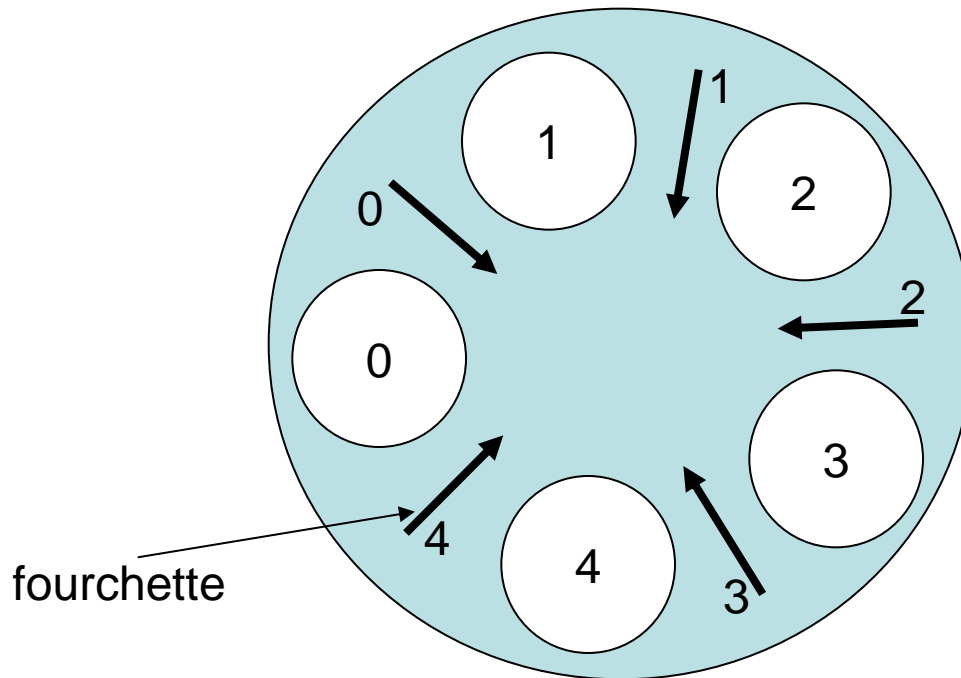
# Le problème des lecteurs-rédacteurs

```
public void startWrite() {  
    wlock.P();  
}
```

```
public void endWrite() {  
    wlock.V();  
}  
}
```

# Le problème des philosophes

Dans ce problème on considère 5 philosophes dont les seules activités sont manger et penser.



# Le problème des philosophes

Avant de commencer à manger, un philosophe doit acquérir deux fourchettes (ressources critiques).

Avant de commencer à penser, un philosophe libère les fourchettes

répéter

penser

pré-protocole // acquérir deux fourchettes

manger

post-protocole // rendre les fourchettes

# Le problème des philosophes

Chaque philosophe est associé à un thread

Les fourchettes sont associées à des ressources et on doit implémenter l'interface

```
interface Ressource {  
    public void acquire(int i);  
    public void release(int i);  
}
```

# Le problème des philosophes

L'implémentation de l'interface Ressource doit satisfaire:

- Un philosophe mange seulement s'il possède deux fourchettes
- Deux philosophes ne possèdent pas la même fourchette simultanément (exclusion mutuelle)
- Pas de d'inter-blocage (freedom from deadlock).
- Pas d'insuffisance de ressource (freedom from starvation)



# Programme de test

```
class Philosopher implements Runnable {  
    int id = 0;  
    Ressource r = null;  
    public Philosopher(int initId, Ressource initr) {  
        id = initId;  
        r = initr;  
    }  
}
```

# Programme de test (suite)

```
public void run() {  
    while(true) {  
        try {  
            System.out.println(" Philosophe " + id + " pense");  
            Thread.sleep(30);  
            System.out.println(" Philosophe " + id + " a faim");  
            r.acquire(id); // acquisition des ressources critiques  
            System.out.println(" Philosophe " + id + " mange");  
            Thread.sleep(40);  
            r.release(id); // libère les ressources  
        } catch (InterruptedException e) { return;}  
    }  
}
```

# 1<sup>ère</sup> tentative

On associe à chaque fourchette (ressource) un sémaphore binaire.  
Le pré-protocole exécuté par le philosophe  $i$  s'écrit:

```
fork[i].P();  
fork[(i + 1) % 5].P();
```

Le post-protocole

```
fork[i].V();  
fork[(i + 1) % 5].V();
```

# 1<sup>ère</sup> tentative

```
class DiningPhilosopher implements Ressource {  
    int n = 0;  
    BinarySemaphore[] fork = null;  
    public DiningPhilosopher(int initN) {  
        n = initN; // nombre de philosophes  
        fork = new BinarySemaphore[n];  
        for (int i = 0; i < n; i++) {  
            fork[i] = new BinarySemaphore(true); // les ressources sont  
                                                    // initialement accessibles  
        }  
    }  
}
```

# 1<sup>ère</sup> tentative (suite)

```
public void acquire(int i) {  
    fork[i].P();  
    fork[(i + 1) % n].P();  
}  
public void release(int i) {  
    fork[i].V();  
    fork[(i + 1) % n].V();  
}  
public static void main(String [] args) {  
    DiningPhilosopher dp = new DiningPhilosopher(5);  
    for(int i = 0; i < 5; i++)  
        new Thread(new Philosopher(i,dp)).start();  
}  
}
```

# Analyse

On montre que deux philosophes ne possèdent jamais la même fourchette

En effet, considérons la fourchette  $i$ .  $\text{fork}[i]$  est un sémaphore binaire et les invariants associés s'écrivent:

$$\text{fork}[i] \geq 0$$

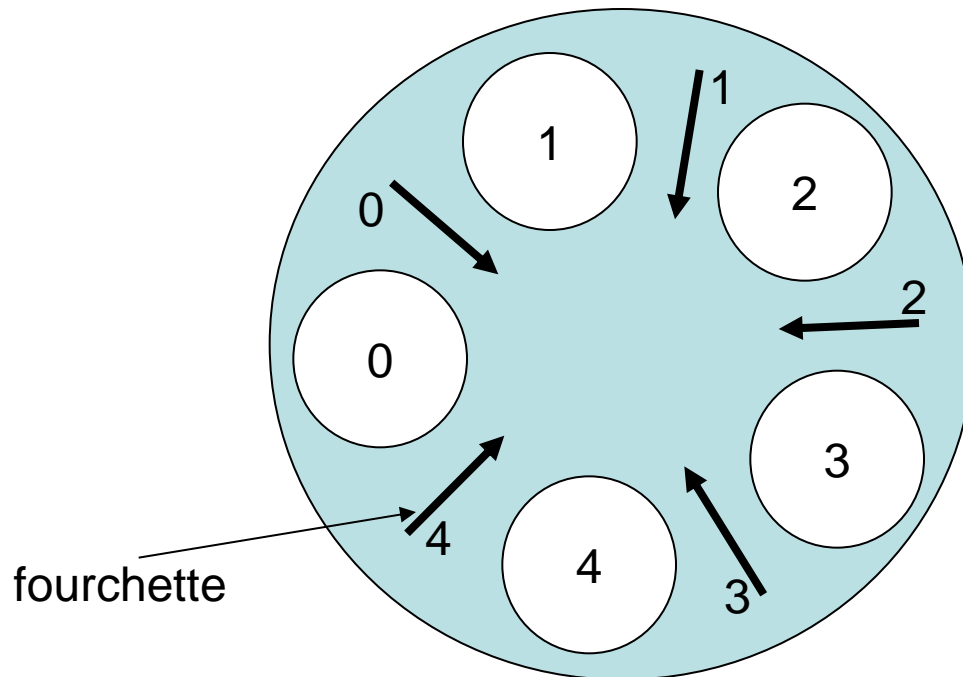
$$\text{fork}[i] = 1 - \# \text{fork}[i].P() + \# \text{fork}[i].V()$$

$$\underbrace{\# \text{fork}[i].P() - \# \text{fork}[i].V()} \leq 1$$

le nombre de philosophes qui disposent de la fourchette

# Analyse 1<sup>ère</sup> tentative (suite)

Malheureusement, il existe une exécution qui conduit à un inter-blocage: Tous les philosophes exécutent `fork[i].P()` séquentiellement.



## 2<sup>ème</sup> tentative

Une solution consiste à empêcher que tous les philosophes puissent exécuter le pré-protocole simultanément. On utilise une sémaphore entière *room* pour limiter le nombre de philosophe à  $n - 1$ .

Le pré-protocole exécuté par le philosophe  $i$  s'écrit:

```
room.P(); CountingSemaphore room = new CountingSemaphore(n-1);  
fork[i].P();  
fork[(i + 1) % 5].P();
```

Le post-protocole

```
fork[i].V();  
fork[(i + 1) % 5].V();  
room.V();
```



# Analyse de la deuxième tentative

La preuve de l'exclusion mutuelle est la même que pour la 1<sup>ère</sup> tentative.

On montre que la deuxième tentative **ne conduit jamais à une absence de ressources (free from starvation)**. Pour cela, on doit supposer que *les processus bloqués dans la file d'attente associée à la sémaphore room sont libérés dans l'ordre d'arrivée*.

Pour les autres sémaphores *fork[i]* on a pas besoin d'une telle hypothèse.

# Analyse de la deuxième tentative

**1<sup>er</sup> cas:** Le philosophe  $i$  est bloqué après l'appel à  $fork[i].P()$ .

Le philosophe  $i-1 \pmod n$  a donc exécuté avec succès  $fork[i].P()$ ,  
comme l'ordre du programme est tel que le philosophe  $i-1$  exécute  
 $fork[i-1].P()$  avant d'exécuter  $fork[i].P()$ ,

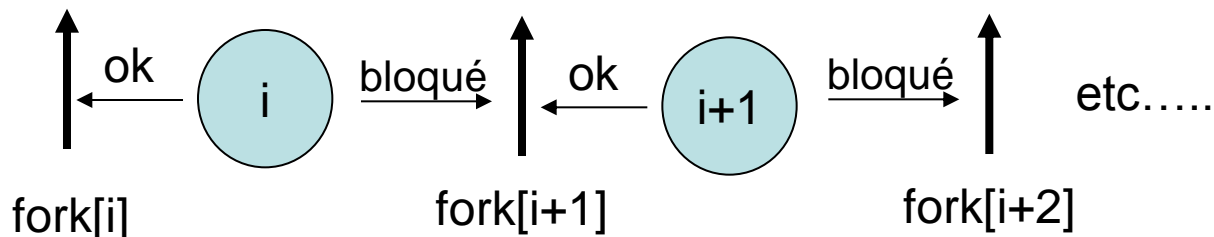
$$fork[i-1].P() \longrightarrow fork[i].P()$$

le processus associé au philosophe  $i-1$  mange... Lorsqu'il aura fini il  
libérera le processus  $i$  qui pourra exécuter  $fork[i+1].P()$ .

# Analyse de la deuxième tentative

**2<sup>ème</sup> cas:** Le processus associé au philosophe  $i$  est bloqué après l'appel à  $fork[i+1].P()$ .

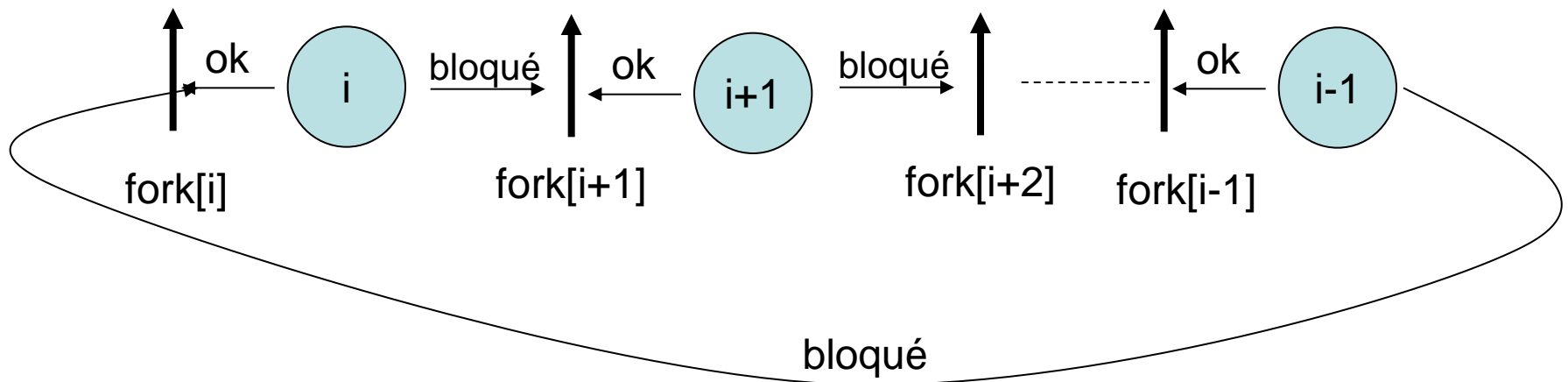
On suppose que le processus  $i+1$  a exécuté  $fork[i].P()$  avec succès et qu'il ne libère jamais cette sémaphore, il y a insuffisance de ressources pour le processus  $i$ . Alors ce processus doit être bloqué sur  $fork[i+2].P()$  sinon il mange et exécutera  $fork[i+1].V()$  libérant le processus  $i$ .



# Analyse de la deuxième tentative

## 2<sup>ème</sup> cas (suite):

Par induction on montre que tous les processus  $j$  doivent être bloqués sur  $fork[j+1].P()$ . C'est une contradiction car dans cette situation tous les processus ont exécutés  $room.P()$  ce qui est impossible



# Analyse de la deuxième tentative

**3<sup>ème</sup> cas:** Le processus associé au philosophe  $i$  est bloqué après l'appel à  $room.P()$ .

On a supposé que cette sémaphore libère les processus selon une stratégie FIFO.

Il doit se trouver  $n-1$  processus dans la section de programme entre  $room.P()$  et  $room.V()$ . L'étude des deux cas précédents montre que ces processus ne peuvent pas rester bloqués dans cette section de code. Alors au moins un processus va exécuter  $room.V()$ , ce qui libérera le processus  $i$ .

# symétrie-asymétrie

Un algorithme distribué est symétrique si au début de l'exécution tous les processus sont dans le même état (variables internes) et que toutes les variables partagées ont la même valeur.

Si on suppose que le système ne dispose pas de mémoire partagée commune ou ne peuvent pas communiquer avec un processus central alors on peut montrer **qu'il n'existe pas de solution au problème des philosophes qui soit symétrique et déterministe.**

Dans la deuxième tentative, la sémaphore *room* est partagée de manière centralisée entre les processus.

# Asymétrie

Une solution asymétrique au problème de philosophes est la suivante.  
Tous les philosophes exécutent l'algorithme précédent sauf le processus  $n-1$  qui exécute

```
public void acquire() {
```

```
    fork[0].P();
```

```
    fork[n-1].P();
```

```
}
```

```
public void release() {
```

```
    fork[0].V();
```

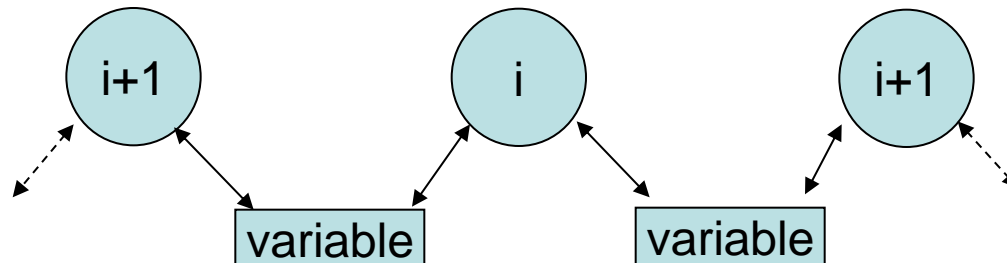
```
    fork[n-1].V();
```



# Impossibilité pour le problème des philosophes

On s'intéresse à l'impossibilité de résoudre le problème des philosophes de manière complètement distribuée (sans processus ou variables communes à tous les philosophes) et symétrique.

Comme le système est complètement distribué les philosophes communiquent uniquement avec leurs voisins, par exemple en partageant une variable commune. Cette variable permet aux philosophes de se mettre d'accord sur lequel va disposer de la fourchette commune.





# Impossibilité pour le problème des philosophes

Les philosophes disposent aussi de variables internes.

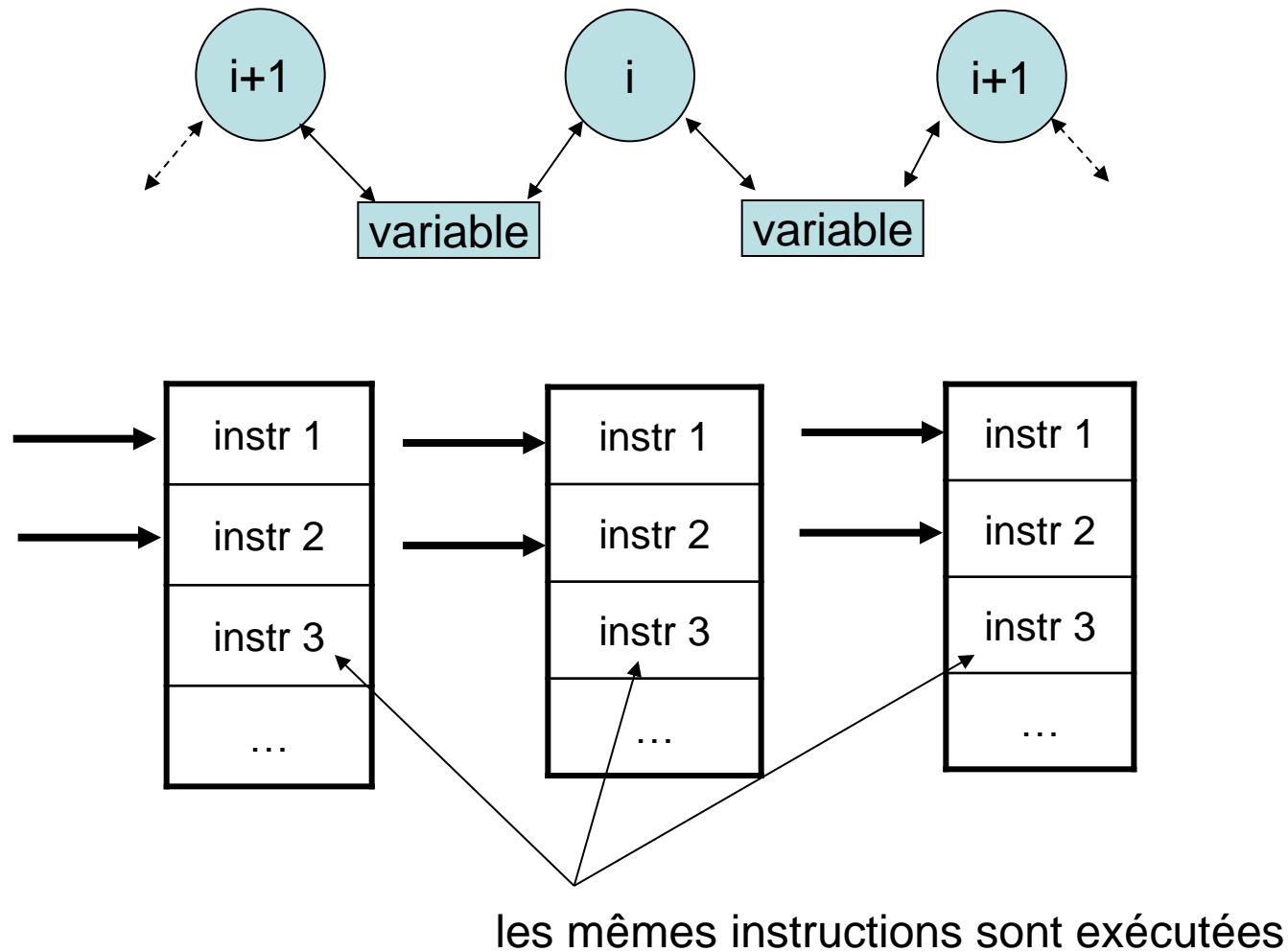
L'algorithme appliqué par chacun des philosophe est déterministe.

La solution au problème est symétrique, c'est-à-dire que toutes les variables (internes et partagées) sont initialisées de manière identique et les programmes exécutés par les philosophes sont les mêmes.

**1<sup>ère</sup> étape:** On montre qu'il existe une séquence d'activation des processus qui fait que les processus sont toujours symétriques.

Supposons que les philosophes soient activés dans l'ordre 1, 2, ..., N et que chaque fois qu'un philosophe est activé il exécute une unique instruction atomique

# Impossibilité pour le problème des philosophes



# Impossibilité pour le problème des philosophes

On a trois cas à traiter

1. l'instruction accède une variable interne
2. l'instruction accède la variable partagée avec le philosophe de droite
3. l'instruction accède la variable partagée avec le philosophe de gauche

On suppose qu'avant l'activation des processus 1, 2, ..., N l'état des processus est symétrique et on montre qu'après que tous les processus ont été activés ils sont dans un état symétrique.

Dans le premier cas, l'opération étant interne aux processus, les processus étant tous dans le même état avant l'exécution de l'instruction, les processus **restent tous dans le même état après l'exécution.**

# Impossibilité pour le problème des philosophes

Dans le deuxième cas, les philosophes accèdent la variable en lecture ou en écriture. Les variable partagées ont toutes la même valeur avant l'exécution par hypothèse.

Donc si c'est une lecture, après exécution de l'instruction les philosophes ont **tous lu la même valeur**.

Si c'est une écriture, les philosophes écrivent la valeur d'une variable interne dans la variable partagée, comme la valeur de cette variable interne est la même pour tous les processus, **la valeur écrite dans la variable partagée est la même pour tous** les philosophes.

Le troisième cas se traite de la même manière.

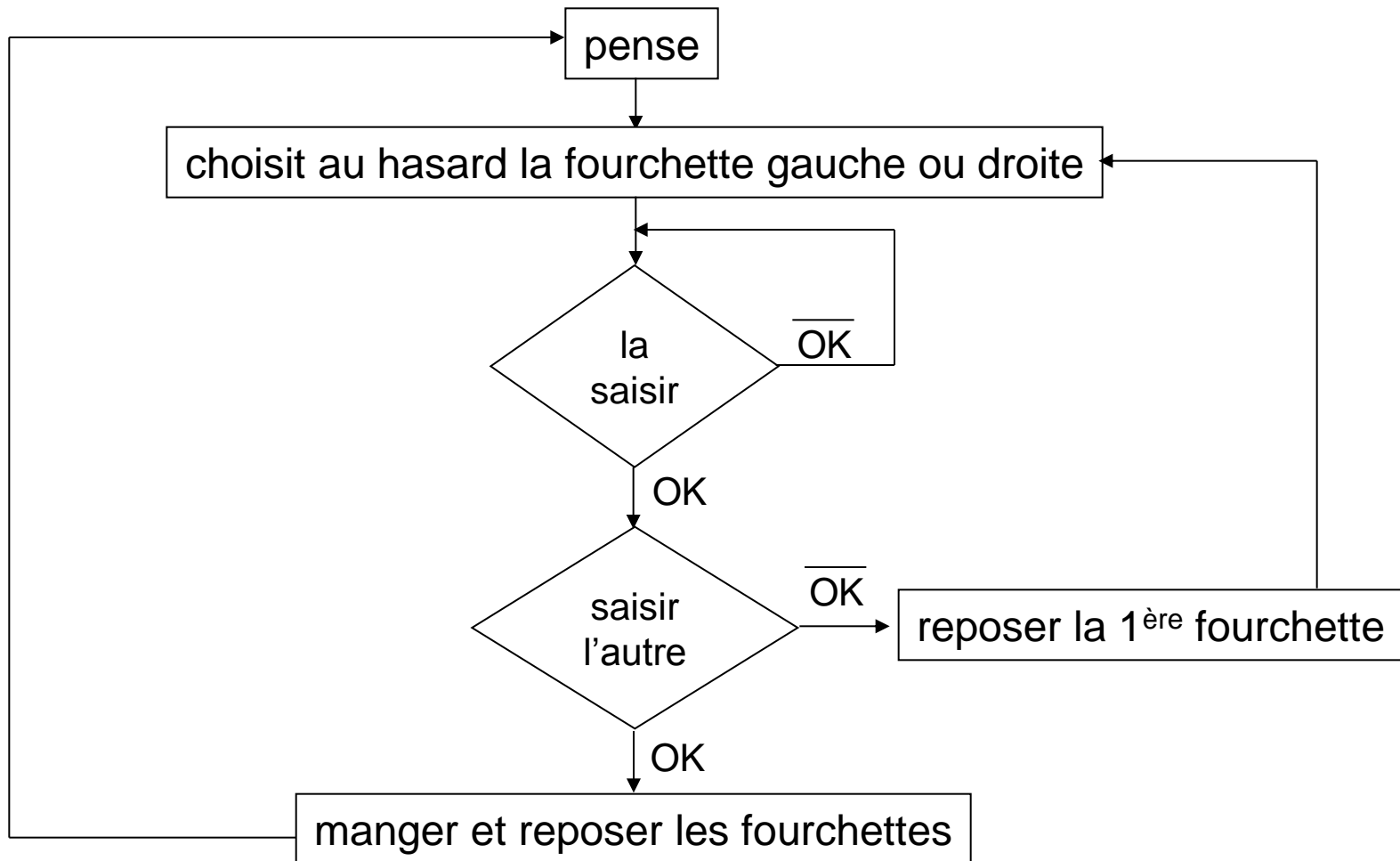
# Impossibilité pour le problème des philosophes

Supposons qu'il existe une exécution d'une instruction après laquelle un philosophe possède ses deux fourchettes. Par symétrie, après l'exécution de la même instruction par le prochain philosophe, ce dernier est en possession de ses deux fourchettes. On poursuit le raisonnement et on montre qu'à la fin de l'activation de tous les philosophes, ils sont tous en possession de deux fourchettes ce qui est impossible.

**On a donc montré qu'il n'existe pas d'algorithme symétrique, complètement distribué et déterministe pour résoudre le problème des philosophes.**

**Il existe des algorithmes probabilistes.**

# Algorithme probabiliste



# Lemme 1

## **Lemme 1:**

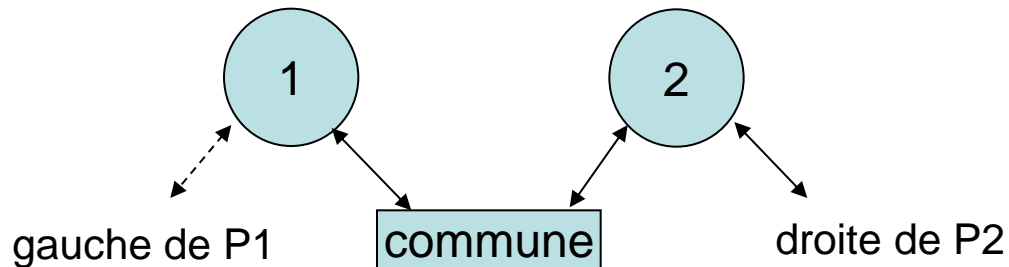
Les penseurs ne peuvent pas être interbloqués. Ils choisissent donc un nombre infini de fois une fourchette (et donc saisissent un nombre infini de fois chacune des fourchettes)

La condition Hold and Wait de la p. 15 du cours n'est pas satisfaite. Les penseurs ne peuvent donc pas être interbloqués. Ils doivent procéder au tirage aléatoire d'une fourchette (et saisir la fourchette correspondante) une infinité de fois.

# Lemme 2

## Lemme 2:

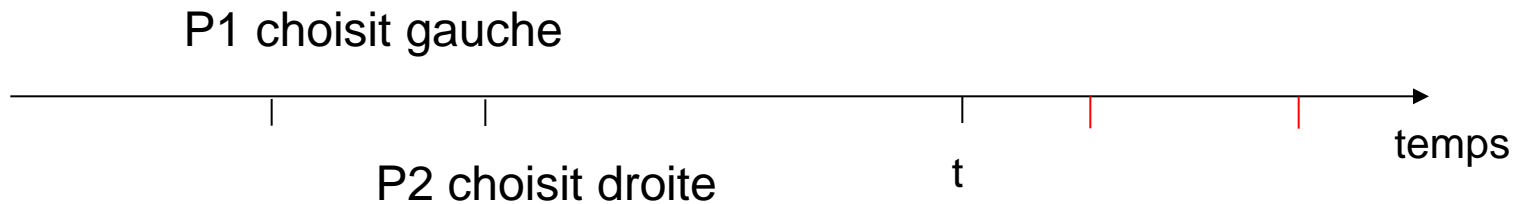
Dans la situation ci-dessous ou P1 saisit la fourchette gauche et P2 la fourchette droite alors P1 ou P2 va manger avec une probabilité positive.





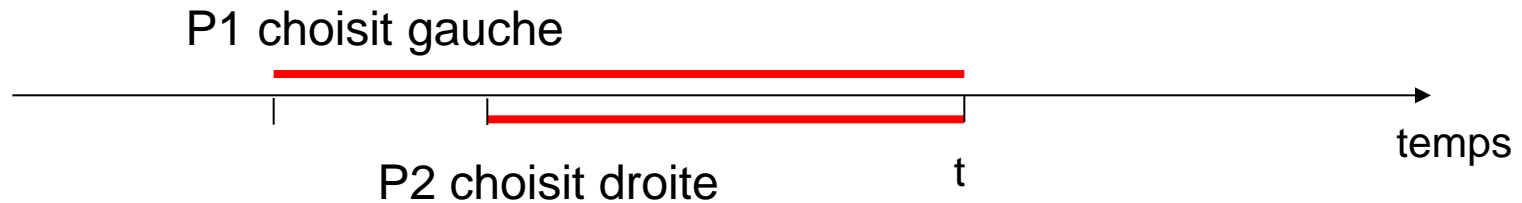
# Lemme 2

Par symétrie on peut supposer que la configuration ci-dessous se produit.



# Lemme 2

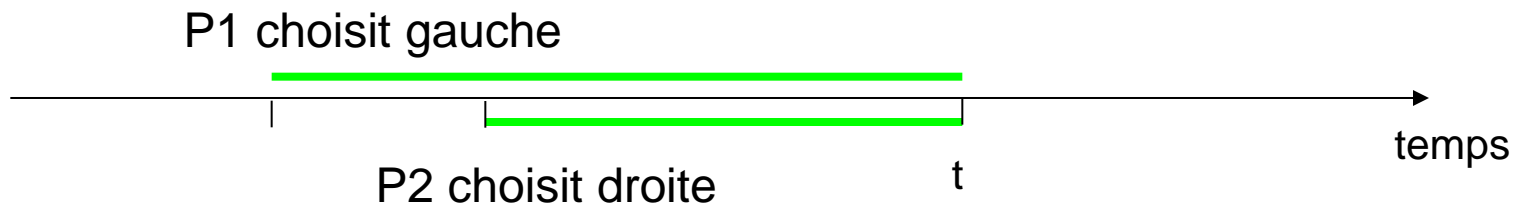
Cas 1: Ni P1 ni P2 essayent de saisir la fourchette commune après le tirage aléatoire et avant  $t$ .



Après  $t$  le premier de P1 ou P2 qui essaye de saisir la fourchette commune va l'obtenir.

# Lemme 2

**Cas 2:** P1 et P2 ont essayé de saisir la fourchette commune après leur dernier tirage aléatoire et avant  $t$ .

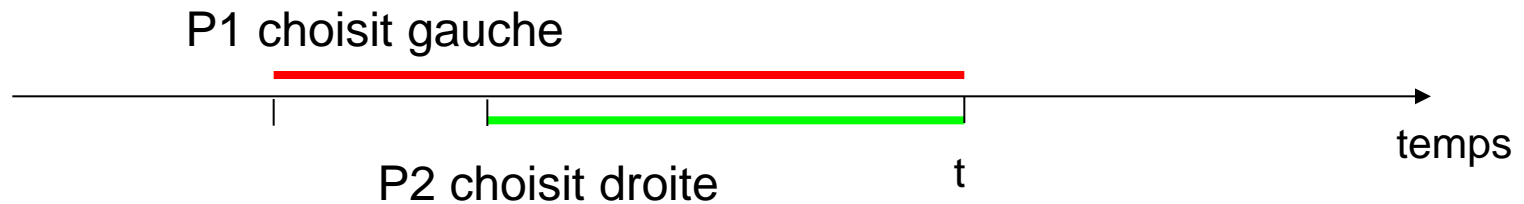


Si le premier à essayer de saisir la fourchette commune la trouve sur la table, il la saisit et commence à manger.

Sinon, il (P1) repose les fourchettes et effectue un nouveau tirage aléatoire mais après  $t$ . Donc l'autre (P2) philosophe va trouver la fourchette commune libre et va commencer à manger.

# Lemme 2

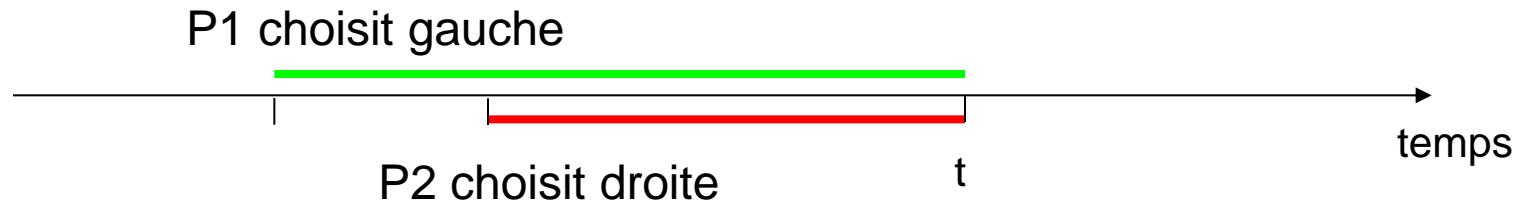
**Cas 3:** P1 ou P2 (exclusif) essaye de saisir la fourchette commune



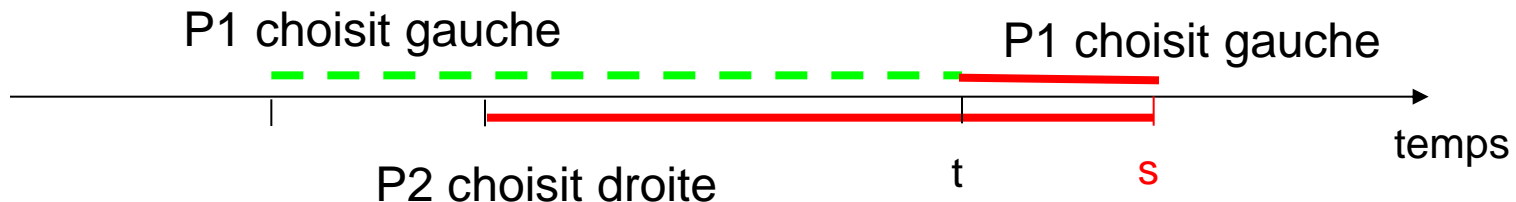
P2 trouve la fourchette commune libre et mange

# Lemme 2

## Cas 4:

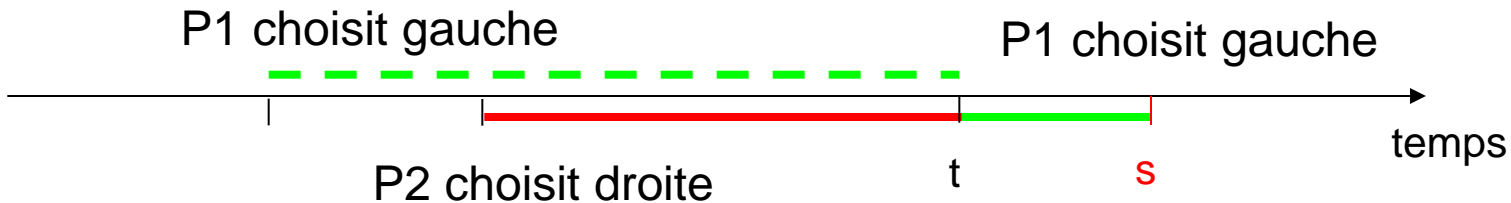


P1 ne trouve pas la fourchette commune sur la table et son prochain tirage aléatoire est à un temps  $s > t$  et *avant que P2 essaye de saisir la fourchette commune*. Avec probabilité  $\frac{1}{2}$  P1 va à nouveau sélectionner la fourchette gauche. On est dans la situation qui correspond au **cas 1**.



# Lemme 2

**Cas 5:** P1 ne trouve pas la fourchette commune sur la table et son prochain tirage aléatoire est à un temps  $s > t$  et *après que P2 essaye de saisir la fourchette commune*.



Dans cette situation, c'est P2 qui commence à manger.

D'après les différents cas considérés, soit P1 soit P2 commence à manger avec un probabilité positive.

# Configuration

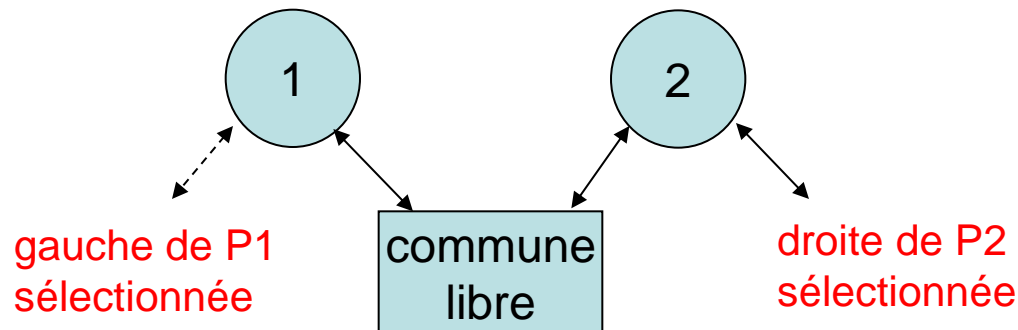
A chaque instant correspond une **configuration** qui consiste en les valeurs des derniers choix effectués par les philosophes.

Des configurations, à des temps différents A et B sont **disjointes** si chaque philosophe a effectué un tirage aléatoire entre A et B.

# Lemme 3

## Lemme 3:

Si au temps  $t$  la configuration est  $A$  alors il arrive avec probabilité 1 une configuration  $B$ , disjointes de  $A$ , et telle qu'il P1 ait choisit sa fourchette gauche (dans  $B$ ) et P2 sa fourchette droite.





# Lemme 3

Depuis la configuration A, chaque philosophe va effectuer un nouveau tirage aléatoire. En effet, il se saisit une infinité de fois d'une fourchette et donc, il effectue une infinité de tirage aléatoire. Il existe donc une configuration G disjointe de A.

Si G satisfait l'énoncé du lemme on a fini.

Sinon, dans G tous les philosophes ont choisis gauche ou tous les philosophes ont choisis droite. Tous les philosophes vont refaire des tirages aléatoires par hypothèse et la probabilité qu'ils sélectionnent toujours soit gauche soit droite est nulle. Avec probabilité 1, on a donc existence d'une configuration disjointes de A qui satisfait l'énoncé du lemme.

# Le théorème

## **Théorème:**

Si l'ordonnanceur assure que toutes les configurations se produisent avec égale probabilités alors les philosophes finissent toujours par manger.

# Insuffisance de ressources

L'algorithme n'assure pas que tous les philosophes mangent à un moment donné (starvation).

On considère deux philosophes P1, P2 adjacents et l'exécution suivante:

1. P1 saisit deux fourchettes et mange
2. P2 test la fourchette commune qui n'est pas disponible
3. P1 restitue les deux fourchettes
4. P1 saisit les deux fourchettes et mange
5. P2 test la fourchette commune qui n'est pas disponible
6. etc.

Dans le cas général on montre qu'il existe un ordonnancement des processus tel que  $n-1$  philosophes ne mangent jamais.

Dans ce cas l'ordonnanceur est supposé un adversaire qui ne laisse pas les configurations se produire au hasard.

# Java en pratique

# Possibilité d'interblocage

On considère l'implémentation suivante d'un objet qui implémente la méthode *swap*

```
class BCell {  
    int value;  
    public synchronized int getValue();  
        return value;  
}  
public synchronized void setValue(int i) {  
    value = i;  
}  
public synchronized void swap(BCell x) {  
    int temp = getValue();  
    setValue(x.getValue());  
    x.setValue(temp);  
}
```

# Possibilité d'interblocage

BCell p,q;

Processus 1

p.swap(q);

Processus 2

q.swap(p);

et l'exécution:

P1.lock(p) // processus 1 obtient le verrou sur p

P2.lock(q) // processus 2 obtient le verrou sur q

q.getValue // processus 1 doit obtenir le verrou, il attend

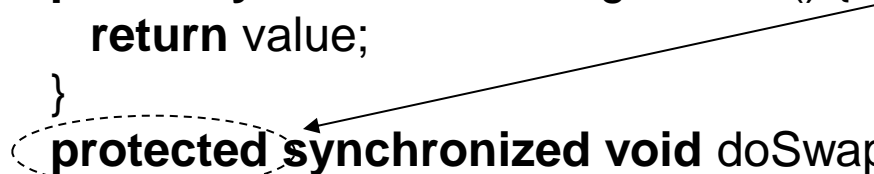
p.getValue // processus 2 doit obtenir le verrou, il attend

La solution consiste à demander les verrous dans le même ordre

# Swap sans interblocage

```
class Cell {  
    int value;  
    public synchronized int getValue() {  
        return value;  
    }  
    protected synchronized void doSwap(Cell x) {  
        int temp = getValue();  
        setValue(x.getValue());  
        x.setValue(temp);  
    }  
    public void swap(Cell x) {  
        if (this == x) return;  
        else if (system.identityHashCode(this) < System.identityHashCode(x))  
            doSwap(x);  
        else  
            x.doSwap(this);  
    }  
}
```

accessible seulement depuis une méthode de la classe



# Data race

Dans un programme concurrent le principal problème est dû aux problèmes de course sur les données qui peut avoir lieu si

*Une variable est accédée par plusieurs threads et au moins l'un deux écrit (modifie le contenu de la variable).*

Pour résoudre ces problèmes, on peut

1. Ne pas partager des variables entre plusieurs threads
2. Ne pas permettre de les modifier
3. Synchroniser les accès à la variable



# Data race

Les problèmes de course surviennent lorsque l'exactitude du résultat de l'exécution des threads dépend de l'ordonnancement (timing) des instructions atomiques.

Exemple: On veut créer un objet uniquement au moment où on doit l'utiliser, l'objet doit être créé une seule fois.

```
Public class LazyInitRace{  
    private ExpensiveObject instance = null;  
    public ExpensiveObject getInstance() {  
        if (instance==null)  
            instance = new ExpensiveObject();  
        return instance;  
    }  
}
```

# Data Race

Dans cet exemple, l'objet peut-être créer plusieurs fois selon l'ordonnancement des instructions. Une solution consiste a synchroniser la méthode getInstance()

```
public synchronized ExpensiveObject getInstance() { ...}
```

# Data Race

C'est le même problème qui survient lorsqu'on exécute *compteur++*;

Pour les manipulations atomiques des données, Java propose des classes adhoc.

AtomicLong

AtomicInteger

AtomicBoolean

....

(voir `java.util.concurrent.atomic`)

# AtomicLong

Par exemple pour implémenter un compteur partagés par plusieurs threads.

```
Public class CountingFactorizer implements Servlet {  
    private final AtomicLong count = new AtomicLong(0);  
  
    public long getCount() {return count.get();}  
  
    public void service(ServletRequest req, ServletResponse resp) {  
        ...  
        count.incrementAndGet(); // ici count++; ne marche pas, service est appelée  
        ...                        // par plusieurs threads  
        ...  
    }  
}
```

# AtomicLong

Une autre solution consiste à synchroniser la méthode service

```
Public synchronized void service(ServletRequest req, ServletResponse resp) {...}
```

D'un point de vue des performances cette solution n'est pas acceptable.

D'une manière générale, on utilise un verrou (lock) chaque fois qu'une opération composée (count++) doit être atomique. Le verrou doit

1. être le même partout où la variable est accédée
2. chaque accès à la variable doit être protégé avec le verrou

# Consistance

Si un ensemble de variables prend des valeurs qui doivent être consistantes alors tous les accès aux variables doivent être protégés par le même verrou.

Les lectures et écritures de variables de type standard sont exécutés atomiquement **sauf** pour les variables *double* ou *long* qui ne sont pas *volatile*. Ces variables sur 64 bits peuvent être accédées comme deux variables de 32 bits. S'il y a plusieurs threads, une lecture peut retourner une valeur qui n'a pas été écrite par un autre thread (*out-of-thin-air*).

Les variables long et double partagées doivent être volatile ou protégées par un verrou.

# Visibilité

Les mécanismes de synchronisation permettent d'exécuter plusieurs instructions de manière atomique. Ce n'est pas la seule fonction de ces mécanismes. La synchronisation permet d'assurer que des modifications (écritures) réalisées sur des variables sont visibles par d'autres threads.

Sans synchronisation, il n'y a pas de garantie qu'une écriture dans une variable soit visible pour un autre processus (écriture en mémoire cache ou dans un registre).

En plus, la synchronisation limite le ré-ordonnancement des instructions.

# Visibilité

```
Public class NoVisibility {  
    private static boolean ready;  
    public static int number;  
    public static class ReaderThread extends Thread {  
        public void run() {  
            while(!ready)  
                Thread.yield();  
            system.out.println(number);  
        }  
    }  
  
    public static void main(String[] args) {  
        new ReaderThread().start();  
        number = 42;  
        ready = true;  
    }  
}
```



# Visibilité

Avec ce programme on peut observer que

1. Le thread `ReaderThread` ne progresse pas, la condition d'attente dans la boucle `while` est toujours valide. Cela peut se produire parce que rien n'oblige l'écriture de la variable `ready` par le thread `main` à être visible.
2. Le thread `ReaderThread` affiche 0 à l'écran. Cela peut se produire si le compilateur ré-ordonne les instructions du thread `main` et l'écriture de `ready` est visible.

Ces comportements sont possibles car le compilateur peut ré-ordonner les instructions pour autant que la sémantique du code pour une exécution séquentielle ne soit pas modifiée et aussi utiliser les registres du processeur comme mémoire cache.

# Visibilité

```
Public class MutableInteger {  
    private int value;  
  
    public int get() {return value;}  
    public void set(int value) {this.value = value; }  
}
```

Un thread qui appelle get() peut ne pas voir la valeur mise à jour par un précédent appel à set(...). Par contre, c'est garanti si on synchronise les routines

```
public synchronized int get() {return value;}  
public synchronized void set(int value) {this.value = value; }
```

# Visibilité - Synchronisation

Synchroniser les routines assure qu'il y a une relation d'ordre entre les appels aux routines, c'est-à-dire l'un apparait après/avant l'autre.

C'est vrai pour les données accédées dans le bloc synchronisé, mais aussi pour toutes les autres variables accédées précédemment dans le code du programme.

# Visibilité - synchronisation

Thread A

y = 1



lock M



x = 1



unlock M



Thread B



lock M



i = x



unlock M



j = y



Toutes les actions avant  
le unlock M sont visibles

# Visibilité - Synchronisation

Les verrous ne permettent pas seulement d'assurer l'exclusion mutuelle lors de l'accès aux variables mais aussi que les modifications des variables soient visibles.

ATTENTION: pour que ca soit vrai, il faut toujours utiliser le même verrou pour protéger l'accès aux variables.

# Visibilité - Volatile

Les variables de type **volatile** assure aussi que les mises-à-jours sont visibles par les autres threads.

Les compilateurs ne sont pas autorisés à ré-ordonner des opérations sur des variables volatiles avec d'autres opérations.

Une lecture d'une variable volatile retourne toujours la valeur correspondant à la dernière mise-à-jour.

Lire une variable volatile assure aussi que toutes les mises-à-jours effectuées par le thread qui a écrit la variable volatile sont visibles. Une écriture à le même effet que libérer un verrou (unlock M) et une lecture le même effet que d'acquérir un verrou (lock M), sans bloquer les processus.

# Visibilité - Volatile

Thread A

$y = 1$



$x = 1$



$M = \dots$



Toutes les actions avant  
l'assignation à  $M$   
sont visibles

lecture de  $M$



$i = x$



$j = y$



# Volatile - exemple

Un exemple typique d'utilisation d'une variable volatile est quand un thread attend qu'un autre thread ait terminé une action en testant une variable de condition.

```
volatile boolean asleep;
```

```
....
```

```
while (!asleep)  
    do something
```

ATTENTION: une variable volatile n'assure pas l'atomicité des accès. Dans l'exemple où on utilise `count++` même si `count` est volatile, ça ne marche pas.



# Volatile

Les situations où une variable volatile est utile sont

1. L'écriture de la variable ne dépend pas de sa valeur actuelle
2. La variable ne doit pas satisfaire des conditions de cohérence avec d'autres variables
3. On n'a pas besoin d'utiliser un verrou pour accéder à la variable.

# Publication

Publier un objet consiste à le rendre accessible depuis des parties du programme qui ne font pas partie de son domaine de visibilité. On peut le faire

1. En utilisant une référence sur l'objet, référence qui possède un plus grand domaine de visibilité que l'objet
2. En utilisant une méthode non privée qui retourne une référence sur l'objet
3. En le passant en argument à une méthode d'une autre classe

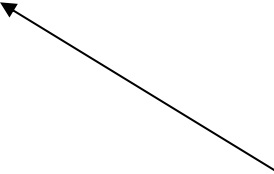
# Publication

Par exemple,

```
public static Set<Secret> knownSecrets;
```

```
Public void initialize() {  
    knownSecrets = new HashSet<Secret>();  
}
```

Peut-être accédé par  
un autre thread alors que  
le constructeur n'a pas terminé  
son exécution



Ou encore

```
class UnsafeStates {  
    private String[] = new String[] { 'A', 'B', ...};  
  
    public String[] getStates() {return states;}  
}
```

# Publication

Un objet se trouve dans un état prévisible et cohérent seulement après que son constructeur ait fini complètement de s'exécuter.

Publier un objet depuis son constructeur est dangereux. Par exemple, en démarrant (`start()`) un thread dans un constructeur et en publiant le pointeur `this` sur l'objet qui se construit.

La publication peut-être explicite, en passant le pointeur `this` en argument ou implicite si l'objet `Runnable` ou le thread exécuté est une sous-classe de l'objet.

# Spécifications

Depuis la version 1.5 de java les comportements des programmes concurrents sont formellement spécifiés.

Les actions inter-threads non synchronisées reconnues par le modèle sont:

- lecture d'une variable partagée (non volatile)
- écriture d'une variable partagée (non volatile)

Les actions inter-threads synchronisées sont:

- lecture d'une variable volatile
- écriture d'une variable volatile
- acquisition d'un verrou (lock)
- libération d'un verrou (unlock)
- la première et la dernière actions d'un thread
- une action qui démarre un thread ou qui détecte la fin d'un thread (`t.isAlive()` ou `t.join()`)

# Synch-order

Toutes les actions de synchronisation sont ordonnées selon un ordre total, c'est-à-dire que les machines virtuelles Java doivent assurer qu'une action est toujours exécutée avant ou après une autre. L'ordre résultant s'appelle synch-order.

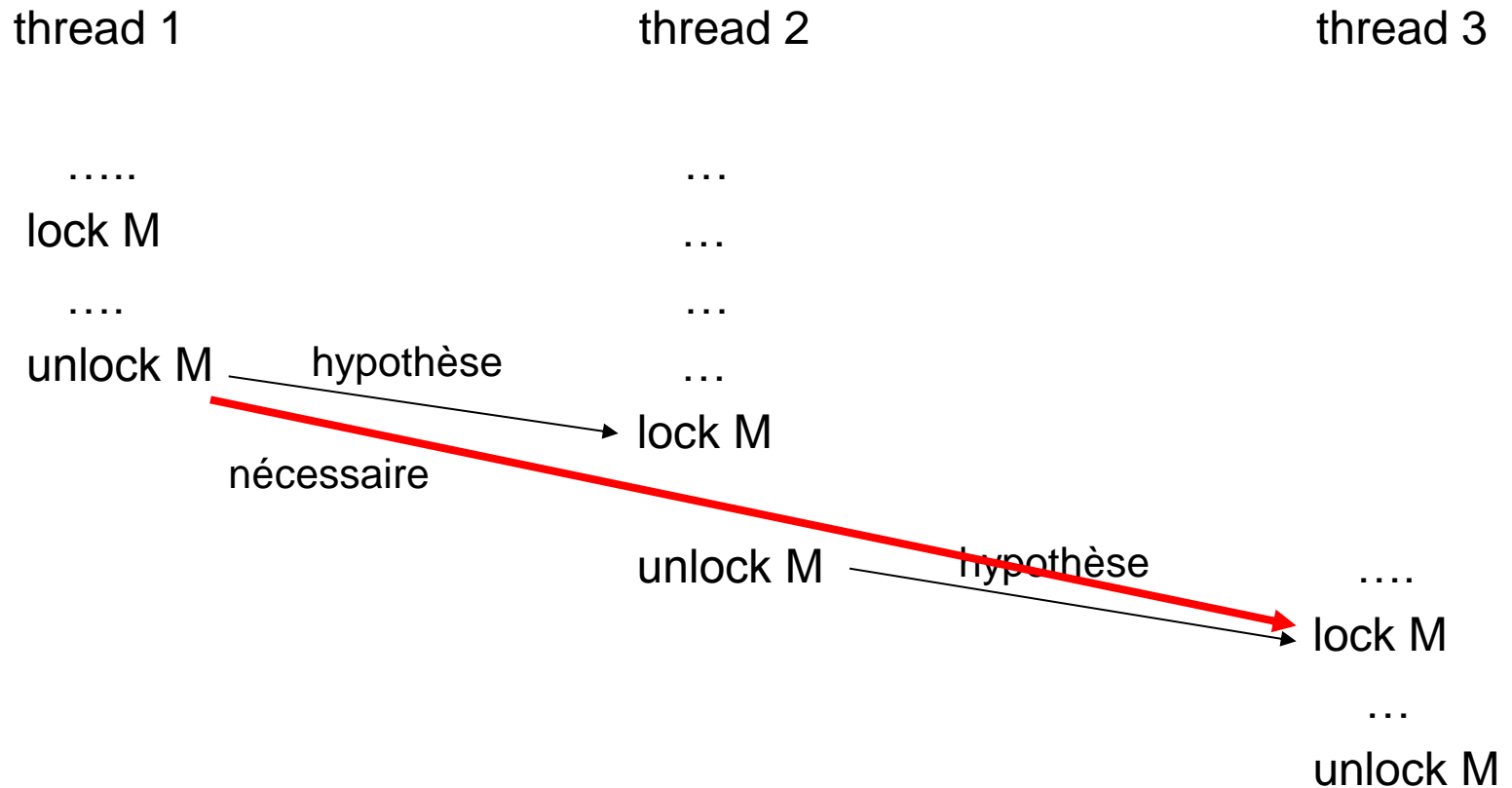
- un `unlock()` sur un verrou est ordonné par rapport aux `lock()` sur le même verrou.
- une écriture d'une variable volatile est ordonnée par rapport à toutes les lectures de la même variable par d'autres threads.
- une action qui démarre un thread apparaît avant la première instruction exécutée par le thread.
- l'écriture des valeurs par défaut (0, false ou null) pour chaque variable apparaît avant la première action de tous les threads

# Synch-order

- la dernière action d'un thread est ordonnée par rapport à toutes les actions d'une autre thread qui détecte que l'exécution du thread est terminée (en utilisant `isAlive` ou `join`)
- si un thread `t1` exécute `t2.interrupt()`, l'interruption est ordonnée par rapport à toutes les actions ou un thread (`t2` ou un autre) détecte l'interruption (en utilisant `isInterrupted`, `interrupted` ou parce qu'une exception `InterruptedException` est levée).

Lorsqu'on raisonne sur les entrelacements possibles, les seuls entrelacements valides sont ceux pour lesquels le synch-order est **total** (ou la relation synch-order peut-être étendue à une relation d'ordre totale)

# Synch-order





# Happen-before

On décrit une relation d'ordre qui s'appelle happen-before. Si une action **a** apparaît avant une action **b** par rapport à cette relation d'ordre, alors **a** est visible pour **b**. On note  $hb(a,b)$  pour indiquer que **a** se produit avant **b**.

- si **a** et **b** sont des actions d'un même thread et que **a** apparaît avant **b** dans l'ordre du programme alors  $hb(a,b)$ .
- si **a** apparaît avant **b** par rapport à l'ordre synchron-order alors  $hb(a,b)$ .
- si  $hb(a,b)$  et  $hb(b,c)$  alors  $hb(a,c)$ .

ATTENTION: la relation  $hb$  sur les actions d'un même thread n'entraîne pas que les actions sont réellement exécutées dans cet ordre, si la sémantique intra-thread n'est pas modifiée le compilateur peut les ordonner différemment. Par contre, les actions sont visibles par les autres threads dans l'ordre.

# Programmes correctement synchronisés

Deux accès à une variable partagée sont conflictuels si au moins un des accès est une écriture .

On a un problème de course sur les données (data-race) si des accès conflictuels ne sont pas ordonnés par la relation happen-before.

Un programme est bien synchronisé si pour tous les entrelacements valides il n'y a pas de problème de course sur les données

Synchronisation  
non  
bloquante  
  
(wait-free)

# Généralités

**But:** développer des mécanismes de synchronisation non bloquants.  
On espère

- Améliorer les performances
- Etre plus tolérant aux pannes en évitant la situation où un processus cesse de s'exécuter tout en possédant un ou plusieurs verrous.
- Eviter les problèmes d'interblocage

# Types de registres

On définit trois types de registres

- sûr (safe)
- régulier (regular)
- atomique (atomic)

Les différents types de registres sont caractérisés par leur comportement lorsque plusieurs processus les écrivent/lisent simultanément.

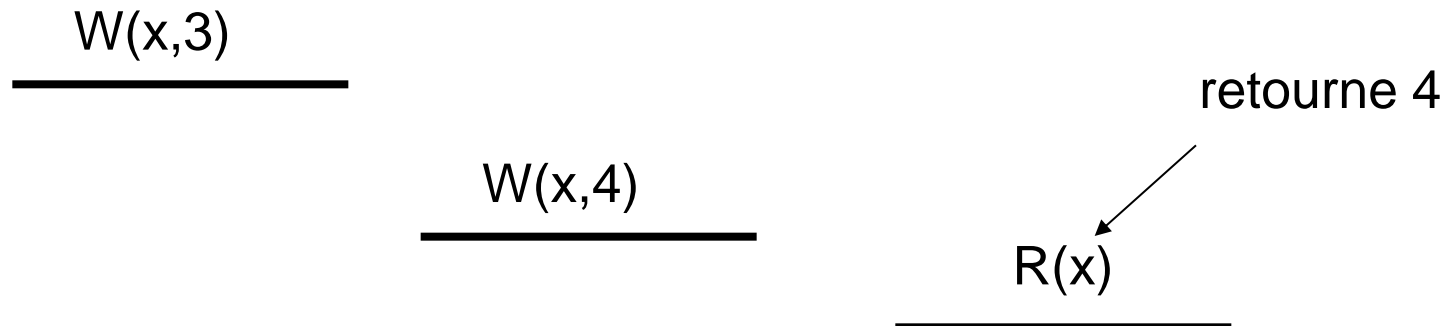
Chacun de ces registres peut être accédés par un seul rédacteur, par plusieurs rédacteurs, par un seul lecteur, par plusieurs lecteurs. Les différents types d'accès permis par les registres sont notés par SW, MW, SR, MR, SWMR,...

# Registres sûrs

Un registre est sûr (safe) si une lecture qui n'est pas simultanée à une écriture retourne la dernière valeur écrite dans le registre.



Dans cette situation le registre est sûr si la lecture  $R(x)$  retourne la valeur 3.



# Registres sûrs

W(x,3)

W(x,4)

R(x)

R(x) peut retourner 3 ou 4, les écritures étant simultanée on ne sait pas laquelle est effective avant l'autre.

# Registres sûrs

W(x,3)

R(x)

R(x) peut retourner n'importe quelle valeur, le comportement du registre n'est pas défini si une écriture et une lecture sont simultanées. De même dans la situation suivante.

W(x,3)

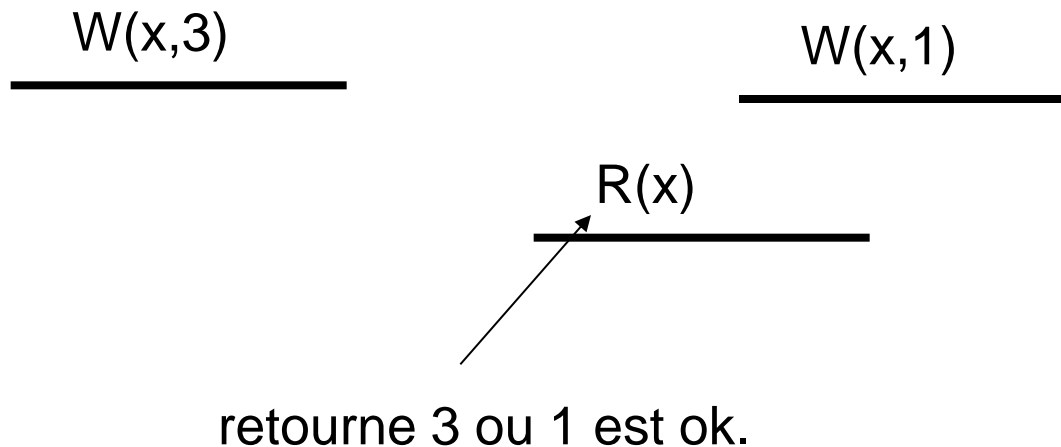
W(x,1)

R(x)

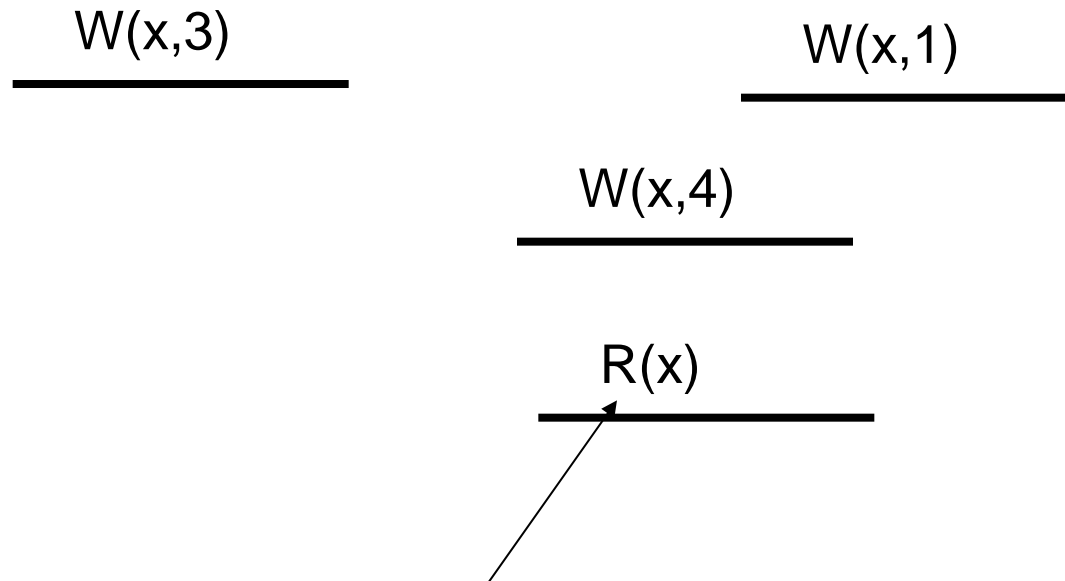


# Registre réguliers

Un registre régulier est sûr et lorsqu'une lecture et plusieurs écritures sont simultanées, il retourne (lecture) soit la dernière valeur écrite, soit une valeur écrite concurremment



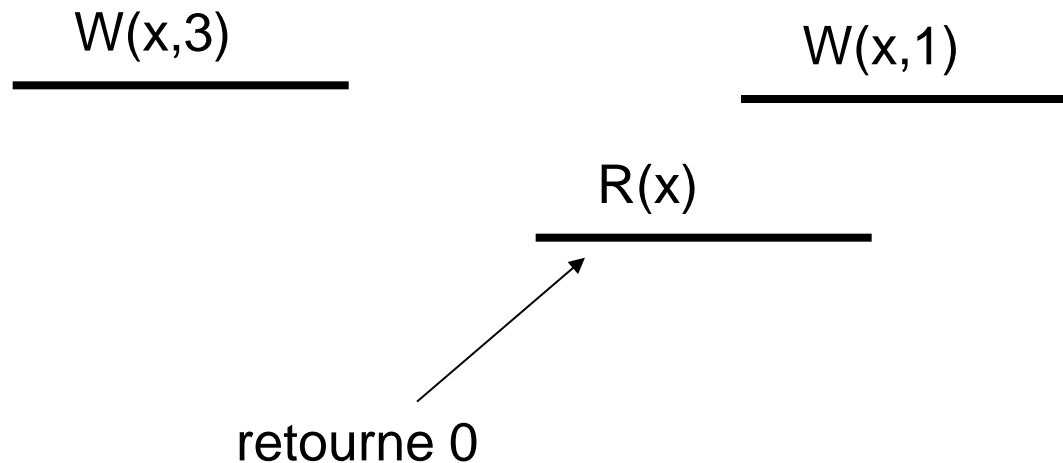
# Registres réguliers



retourne 4, 3 ou 1

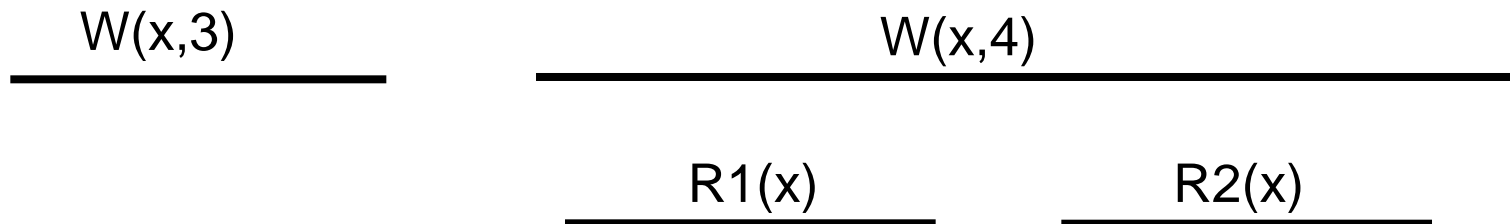
# Registres réguliers

L'exemple ci-dessous montre une exécution où le registre n'est pas régulier, la valeur retournée ne correspond pas ni à la dernière écriture ni à l'écriture concurrente.



# Registres atomiques

Un registre est atomique s'il est régulier et linéarisable. La condition de linéarisabilité détermine le comportement du registre lorsque plusieurs lectures sont séquentiellement effectuées. Le registre est linéarisable si des lectures successives sont bien vues successives par le registre.



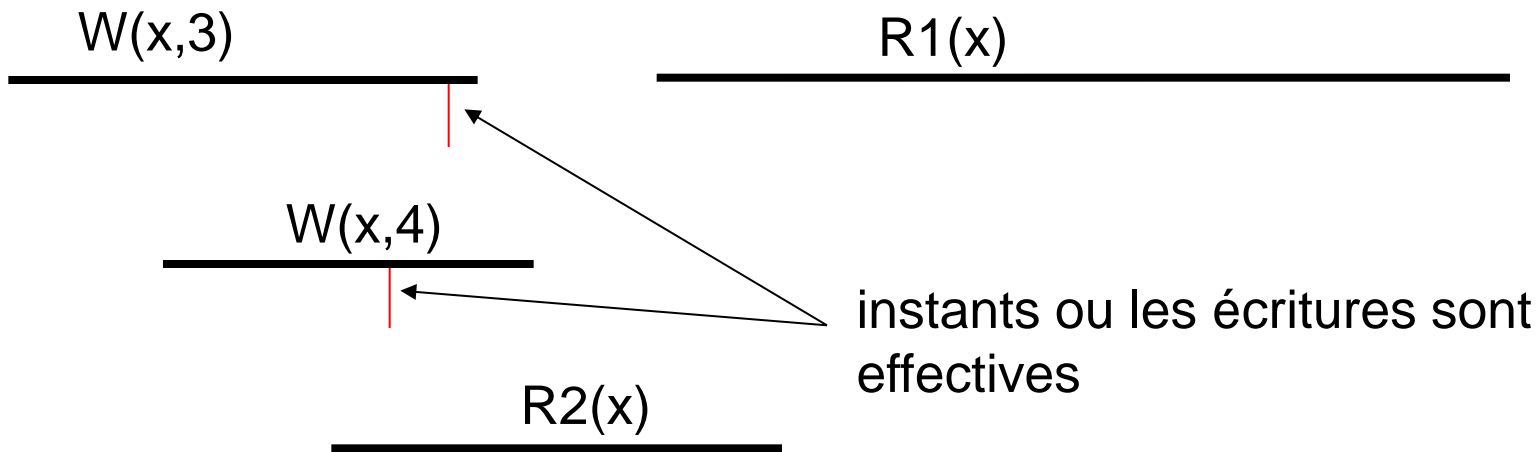
R1 retourne 3 et R2 retourne 3, ok

R1 retourne 3 et R2 retourne 4, ok

R1 retourne 4 et R2 retourne 3, pas atomique, mais régulier

R1 retourne 4 et R2 retourne 4, ok

# Registres atomiques



$R2$  retourne 4 et  $R1$  retourne 3 est atomique. Correspond à la situation où l'écriture de 3 est effective après l'écriture de la valeur 4.

# Registres atomiques

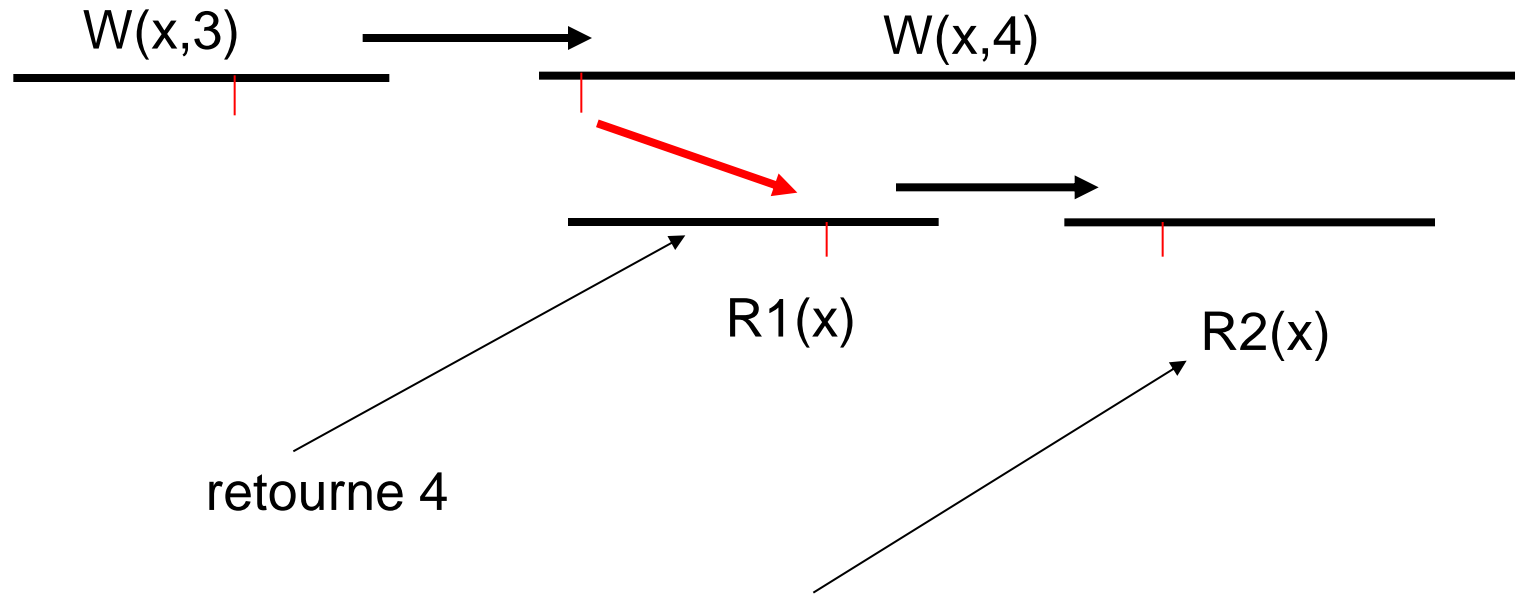
Pour définir la notion de registre atomique plus formellement, on introduit une notion d'ordre de précedence.

Les lectures/écritures d'un même processus sont toutes ordonnées par une relation de précedence décrite par l'ordre séquentiel du programme.

On imagine que les accès aux registres sont effectifs à un instant donné instantanément.

Une exécution particulière détermine un ordre inter-processus qui doit être **total**.

# Exemple



R2 ne peut retourner que 4 pour assurer que l'ordre soit total

# Définitions

On note  $v^i, i = 1, 2, \dots$  les différentes valeurs prises par un registre pendant une exécution donnée.

A chaque valeur correspond *une unique opération d'écriture*, on note ces opérations  $W^i, i=1, 2, \dots$  Tapez une équation ici.

On note  $R^i, i = 1, 2, \dots$  les lectures du registre qui retournent les valeurs  $v^i, i = 1, 2, \dots$

Une exécution particulière contient un unique  $W^i$  mais peut contenir plusieurs  $R^i$

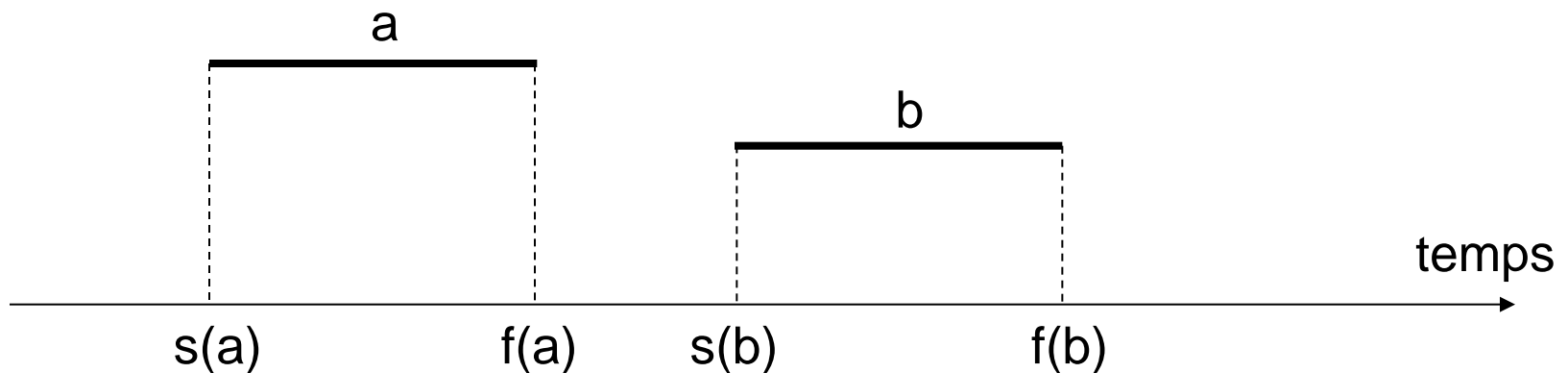


# Définitions

A chaque opération  $a$  de lecture ou écriture on associe un temps de début  $s(a)$  et de fin  $f(a)$ . Pour deux opérations  $a$  et  $b$  on a

$$a \rightarrow b$$

si  $f(a) < s(b)$



# Définitions

Un registre est régulier si:

- Une lecture ne précède jamais une écriture, c'est-à-dire on a jamais  $R^i \rightarrow W^i$
- Une lecture ne retourne jamais une valeur passée qui à été modifiée, on a jamais  $W^i \rightarrow W^j \rightarrow R^i$

Un registre atomique satisfait en plus

$$\text{si } R^i \rightarrow R^j \text{ alors } i \leq j$$

Une registre est sûr s'il est atomique lorsqu'on se restreint aux exécutions sans chevauchement (overlap)

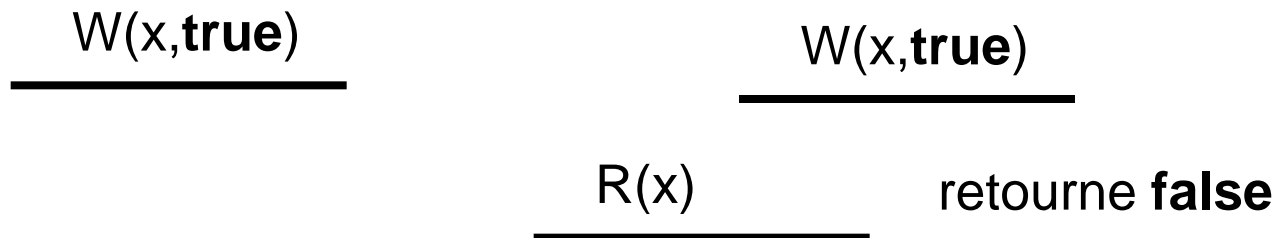
# Registre Java sûr

Le prototype du registre SRSW booléen sûr en Java est donné par le code suivant

```
class SafeBoolean {  
    boolean value;  
    public boolean getValue() {  
        return value;  
    }  
    public void setValue(boolean b) {  
        value = b;  
    }  
}
```

# Registre Java régulier

La seule situation où le registre précédent n'est pas régulier est la situation où deux écritures consécutives sont identiques (soit **false** soit **true**) et qu'une lecture concurrente retourne la valeur complémentaire.



Pour éviter cette situation, on ne réécrit pas la valeur du registre.

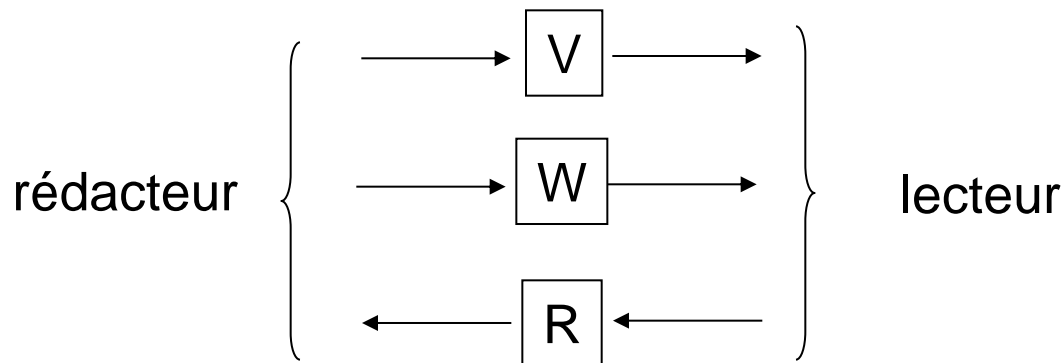
# Registre Java régulier

```
class RegularBoolean {  
    boolean prev; // on mémorise la valeur précédente, n'est pas partagée  
    SafeBoolean value; // un registre safe  
    public boolean getValue() {  
        return value.getValue();  
    }  
    public void setValue(boolean b) {  
        if (prev != b) {  
            value.setValue(b);  
            prev = b;  
        }  
    }  
}
```

# Registre atomique

Pour implémenter un registre atomique, on note la méthode **setValue()** par **change()** pour indiquer que l'écriture est effective seulement si la valeur est différente

Pour construire le registre V atomique on utilise deux registres W et R modifiés par le rédacteur et le lecteur respectivement (**registres réguliers**)



# registre atomique

Protocole rédacteur

change V

**if** W == R **then** change W

Protocole lecteur

1. **if** W == R **then return** v

2. read x := V

3. **if** W <> R **then** change R

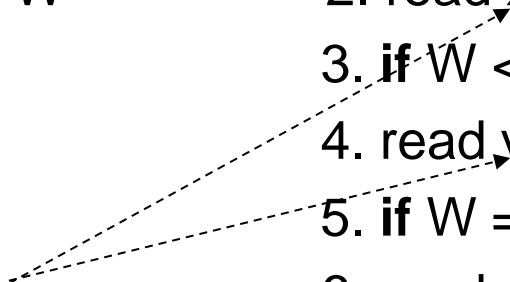
4. read y := V

5. **if** W == R **then return** v

6. read v := V

7. **return** x

variable locales



# Preuve

On montre pour commencer que l'on a jamais  $R^i \rightarrow W^i$

On procède par l'absurde, soit  $r$  la (sous-)lecture qui retourne la valeur erronée. Par définition, on a

$$f(r) < f(R) < s(W)$$

Le registre  $V$  est régulier, alors aucune lecture  $r$  ne peut satisfaire  $f(r) < s(W)$ .



# Preuve

On montre que le registre est régulier, en conjonction avec ce qu'on a déjà prouvé, il reste à voir qu'on n'a jamais  $W^i \rightarrow W^j(w) \rightarrow R^i(r)$

On procède par l'absurde. On note par  $r$  la (sous-)lecture du registre  $V$  qui retourne la valeur dans  $R^i$ , et  $w$  l'écriture dans  $W^j$ .

Le registre  $V$  est régulier, alors on doit avoir  $s(r) < f(w)$ . C'est possible si la lecture se termine à la ligne 1. du protocole du lecteur et retourne la valeur de la variable locale  $v$ .

Mais cela ne peut pas se produire sous les hypothèses puisque le protocole d'écriture modifie la valeur de  $W$  avant de terminer.

En fait, on a démontré que

$$W \rightarrow R \Rightarrow w \rightarrow r$$

# Preuve

Il reste à montrer que le registre est atomique, c'est-à-dire que  
si  $R^i \rightarrow R^j$  alors  $i \leq j$ , c'est-à-dire  $\neg (W^j \rightarrow W^i)$

On procède par l'absurde

soient  $r_i, r_j$  les sous-lectures du registre  $V$ . On note  $r_i \in R^i$   
pour indiquer que la sous-lecture a bien été effectuée pendant la  
lecture  $R^i$

ce qui est le cas si le lecteur quitte le protocole aux lignes 5 ou 7. S'il  
quitte le protocole à la ligne 1, alors la sous-lecture a été effectuée  
avant.

Pour les écritures on a pas d'ambiguïté et on a bien  $w_j \rightarrow w_j$

On suppose  $i \neq j$  et on a alors  $r_i \rightarrow r_j$  (on a bien deux lectures  
qui retournent des valeurs différentes)

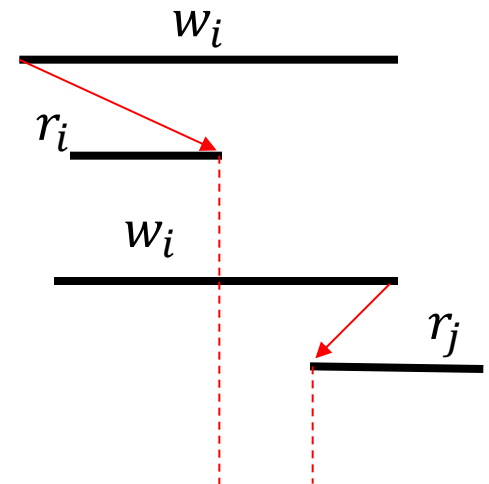
# Preuve

En tenant compte du fait que  $V$  est régulier, on doit nécessairement avoir

$$s(w_i) < f(r_i) \text{ car } \neg(r_i \rightarrow w_i)$$

et aussi

$$s(r_j) < f(w_i) \text{ car } \neg(w_j \rightarrow w_i \rightarrow r_j)$$



comme  $r_i \rightarrow r_j$ , on a  $s(w_i) < f(r_i) < s(r_j) < f(w_i)$

et la valeur de  $W$  ne change pas pendant l'intervalle  $[f(r_i), s(r_j)]$  (\*)

# Preuve

La lecture  $r_i$  peut s'exécuter aux lignes 2, 4 ou 6 du protocole.

**cas 1:** read  $x := V$  ligne 2. Alors  $R_i$  retourne la valeur de  $x$  à la ligne 7. de son protocole. On voit dans l'exécution du protocole qu'entre la ligne 2. et la ligne 7. la valeur de  $W$  doit nécessairement changer, ce qui contredit la remarque (\*) ( on a  $R^i \rightarrow R^j$  ).

**cas 2:** read  $v := V$  ligne 4. Alors  $R_i$  retourne à la ligne 5 de son protocole après avoir testé  $W == R$ . Comme la valeur de  $W$  ne change pas dans l'intervalle  $[f(r_i), s(r_j)]$  l'exécution du protocole pour  $R^j$  doit se terminer à la ligne 1. et les valeurs retournées sont les mêmes, une contradiction.

**cas 3:** read  $v := V$  ligne 6.  $R^i$  retourne à la ligne 1. de son protocole après avoir testé  $W == R$  et on obtient la même contradiction qu'as cas 2.

# Registres multivalués SRSW

Pour implémenter un registre multivalué SRSW atomique on utilise un tableau de registres **booléens atomique SRSW**.

L'idée est de représenter un nombre compris dans l'intervalle **0 ... maxVal-1** par un bit positionné à **true** à la position correspondante dans le tableau.

Par exemple, on code la valeur 3 par le tableau:

0	1	2	3	4
false	false	false	true	false

# Registres multivalués SRSW

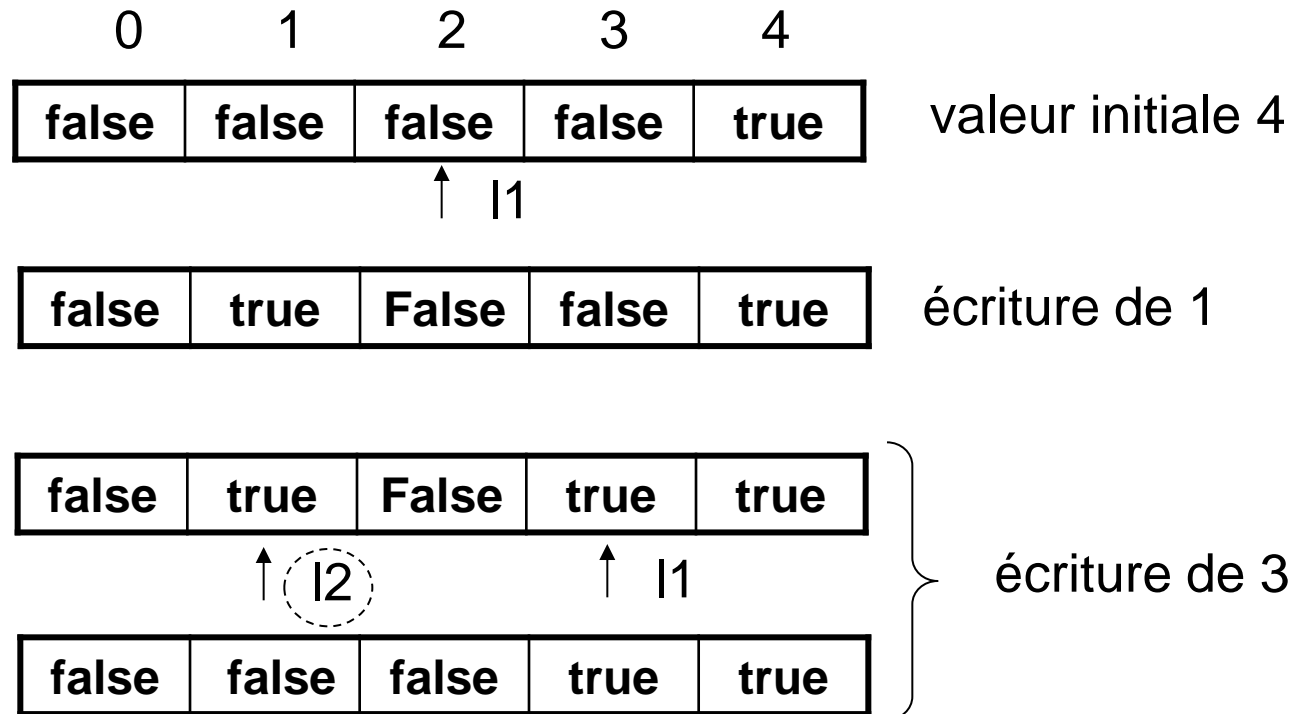
Pour l'écriture la procédure est de la forme:

```
public void setValue(int x) {  
    A[x] = true;                // on dépose la valeur  
    for( int i = x - 1; i >= 0; i--)  
        A[i] = false;          // on supprime la valeur précédente si <  
}
```

Il est nécessaire de positionner l'entrée correspondante dans un premier temps et de supprimer les autres entrées. Sinon, un lecteur pourrait trouver toutes les entrées à **false**.

# Registres multivalués MRSW

Pour le lecteur, on ne peut pas se contenter de parcourir le tableau dans l'ordre croissant des indices pour que le registre soit atomique. En effet, on considère l'exécution où on écrit 4 puis 1 puis 3 dans le registre

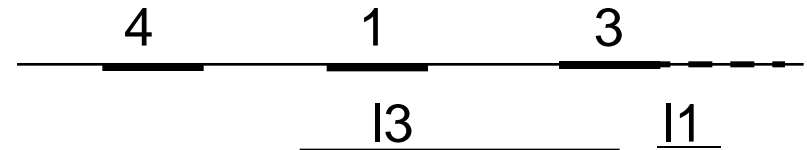


I2, lecture 2 après la première lecture I1

# Registres multivalués SRSW

On a les écritures suivantes  $4 \rightarrow 1 \rightarrow 3$

La première lecture retourne 3



La deuxième lecture retourne 1

Le registre n'est pas atomique puisque la seconde lecture retourne la valeur précédant l'écriture concurrente et la première lecture la valeur écrite concurremment.

Le registre est régulier. En fait c'est un registre MRSW régulier (même si les registres SRSW sont atomiques).



# Registres multivalués atomique SRSW

```
class MultiValued {  
    int n = 0;  
    boolean A[] = null;  
    public MultiValued(int maxVal, int initVal) {  
        n = maxVal; A = new boolean [n];  
        for(int i = 0; i < n; i++) A[i] = false;  
        A[initVal] = true;  
    }  
    public void setValue(int x) {  
        A[x] = true;                // on dépose la valeur  
        for( int i = x - 1; i >= 0; i--)  
            A[i] = false;           // on supprime la valeur précédente si <  
    }
```

# Registres multivalués atomique SRSW

```
public int getValue() {  
    int j = 0;  
    while(!A[j]) j++;           // forward scan  
    int v = j;  
    for(int i = j-1; i >= 0; i--)  
        if (A[i]) v = i;       } backward scan  
    return v;  
}
```

Cette construction est aussi valable pour plusieurs lecteurs, on a un registre MRSW atomique.

# registre MRSW atomique

Pour permettre plusieurs lecteur d'accéder simultanément le registre, l'idée est de maintenir un sous-registre par lecteur, c'est-à-dire utiliser un tableau de SRSW registres  $V[n]$  (des registres **réguliers**)

Le rédacteur écrit les nouvelles valeurs dans tous les sous-registres.

Le registre ainsi obtenu n'est pas atomique.

Il faut qu'un lecteur s'assure après la lecture de son registre qu'il a obtenu la dernière valeur écrite.

Pour cela, les registres SRSW de base sont constitués de deux champs, **value** et **ts** pour la valeur et un **timestamp**.

Un tableau **Comm[i][j]** indique la valeur lue par le lecteur i au lecteur j.

# registre MRSW atomique

```
class MRSW {  
    int n = 0;  
    SRSW V[] = null; // tableau de registres  
    SRSW Comm[][] = null; // communication inter lecteur  
    int seqNo = 0;  
    public MRSW(int readers, int initVal) {  
        n = readers;  
        V = new SRSW[n];  
        for(int i = 0; i < n; i++)  
            V[i].setValue(initVal,0); // initialisation de la valeur et timestamp = 0  
        Comm = new SRSW[n][n];  
        for(int i = 0; i < n; i++) for(int j = 0, j < n; j++)  
            Comm[i][j] .setValue(initVal,0);  
    }  
}
```

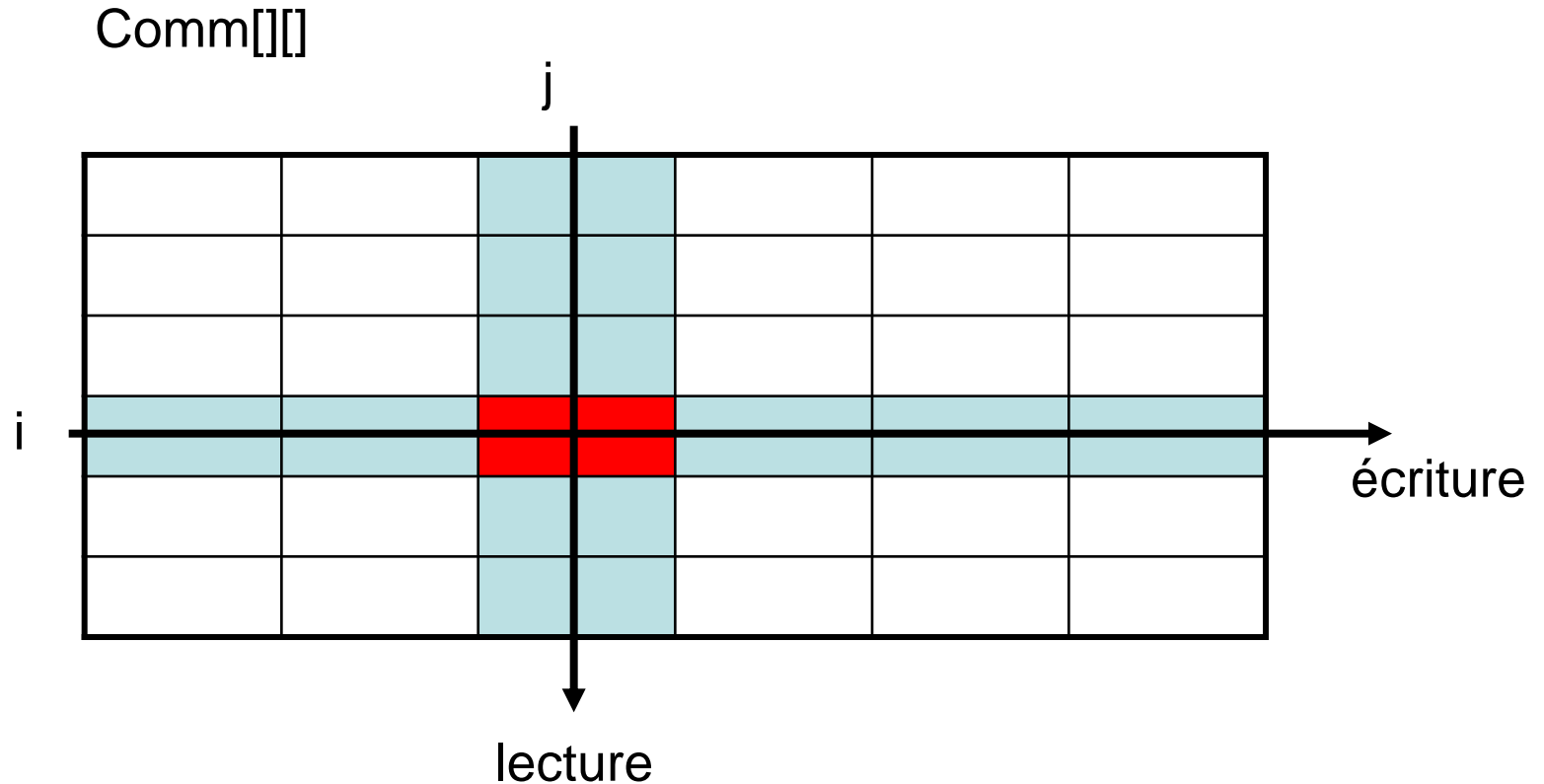
# Registre MRSW atomique

```
public int getValue(int r) { // lecteur r lit son propre registre
    SRSW tsv = V[r]; // variable locale
    for(int i = 0; i < n; i++)
        if (Comm[i][r].getTS() > tsv.getTS())
            tsv = Comm[i][r];
    for(int i = 0; i < n; i++) {
        Comm[r][i].setValue(tsv);
    }
    return tsv.getValue();
}
```

# Registre MRSW atomique

```
public void setValue(int x){  
    seqNo++; // numéro de séquence suivant  
    for(int i = 0; i < n; i++)  
        V[i].setValue(x,seqNo);  
    }  
}
```

# Registre MRSW atomique



*si  $R^i \rightarrow R^j$  alors  $\neg (W^i \rightarrow W^j)$*

# Remarque

L'intérêt de travailler avec des objets atomiques (linéarisables) est que n'importe quelle exécution peut s'interpréter comme un entrelacement d'exécutions atomiques.

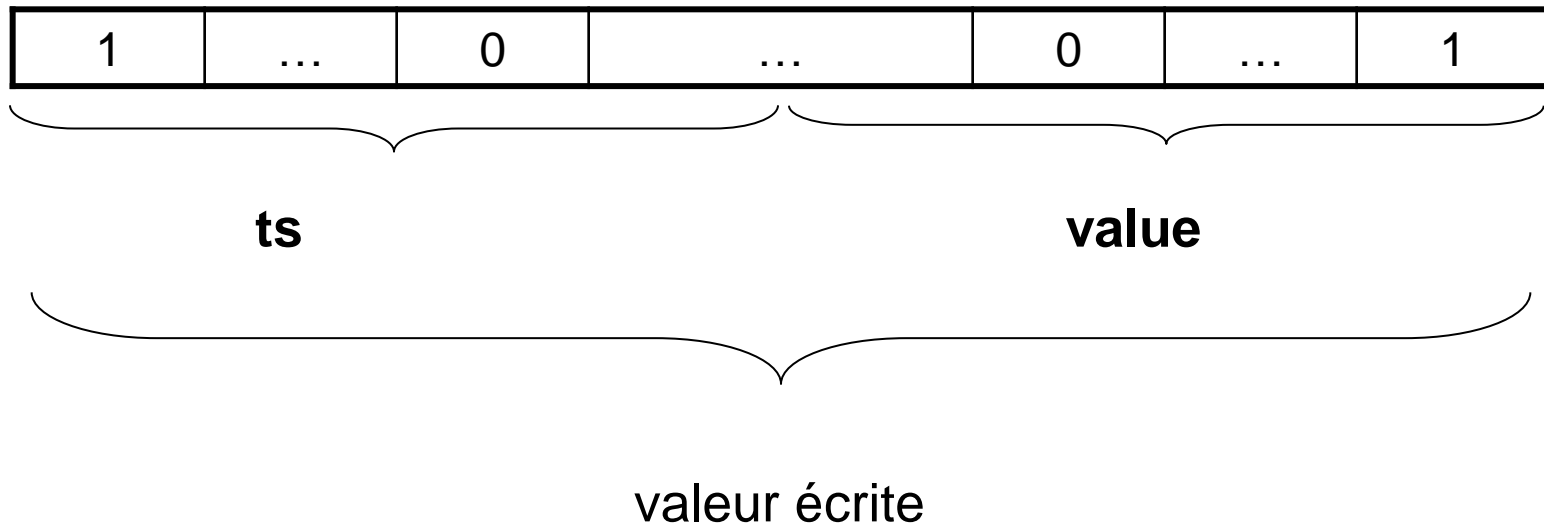
Pour spécifier la sémantique d'un objet on doit spécifier le préconditions admissibles et les postconditions correspondantes. (La documentation augmente linéairement avec le nombre d'objets)

La propriété de linéarisabilité est locale, c'est-à-dire que si tous les objets d'un programme le sont, le programme est linéarisable aussi.



# Remarque

Pour maintenir un timestamp **ts** ainsi qu'une valeur **value** on peut décomposer la valeur écrite dans le registre multivalué atomique comme:



# Autres solutions

La version 1.5 propose le paquetage `java.util.concurrent.atomic`, qui contient, entre autres, la classe

**`AtomicStampedReference<V>`**

Une autre solution utilise un tableau de registres multivalués atomiques.

Cette solution a d'autres applications et est purement algorithmique.

# ThreadLocal

Pour la programmation séquentielle, Java propose des objets de types statiques et non statiques. Les objets non statiques sont répliqués dans chaque instances de la classe, les objets statiques sont communs à toutes les instances de la classe.

Un objet de type **ThreadLocal** est partagé par toutes les instances d'une classe **d'un même thread** et sont différents pour des threads différents.

# Consensus

Le problème du consensus est le suivant:

- i) A chaque processus est assigné une valeur initiale,
- ii) Les processus doivent se mettre d'accord sur une valeur commune

Cette valeur commune doit appartenir à l'ensemble des valeurs initiales pour éviter une solution triviale qui consiste à toujours choisir une même valeur. L'algorithme doit être *wait-free*.

Un objet consensus doit implémenter l'interface suivante

```
public interface consensus {  
    public void propose(int pid, int value);  
    public int decide(int pid);  
}
```

# Consensus – registres atomiques

Le problème du consensus peut-être résolu pour un seul processus avec des registres atomiques.

**Pour deux processus, il n'y a pas de solution wait-free au problème du consensus.**

Pour deux processus, le protocole doit donc permettre aux processus de se mettre d'accord sur une des deux valeurs initiales.

Le protocole est dans un *état bivalent* à un instant donné si les deux valeurs initiales peuvent être choisies selon l'exécution à venir.

Un état bivalent est *critique* si quelque soit la prochaine action exécutée l'état du protocole est non bivalent.

# Etat initial

## **L'état initial du protocole est bivalent:**

On suppose que les deux valeurs initiales sont différentes, P0 obtient 0 et P1 1.

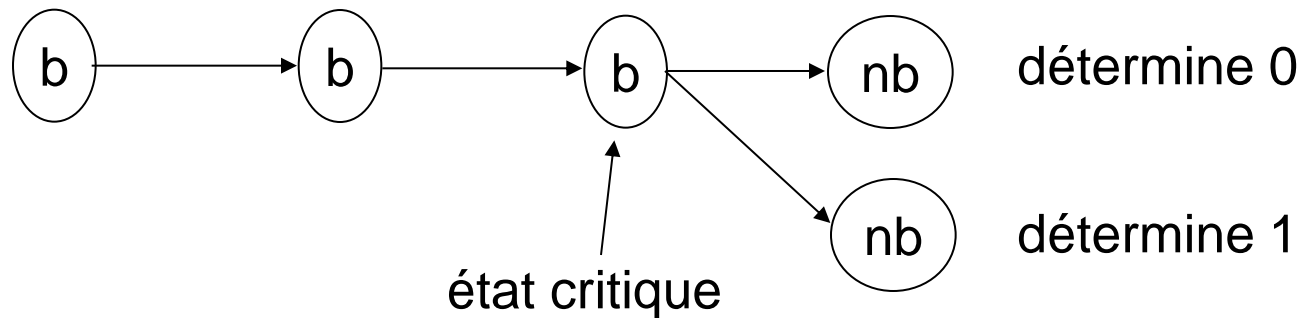
Une première exécution consiste en l'exécution complète du protocole par P0. Comme le protocole est wait-free, le protocole doit choisir la valeur 0 après un nombre fini d'étapes. Ensuite, le protocole est exécuté par P1 qui doit, lui aussi, choisir la valeur 0.

La deuxième exécution consiste en l'exécution du protocole par P1. Cette exécution doit se terminer par le choix de la valeur 1.

Ces deux exécutions montrent que l'état initial est bivalent.

# Etats bivalents

On considère l'exécution du protocole par les deux processus. Comme le protocole est wait-free, il arrive un instant où le protocole passe d'un état bivalent à un état non-bivalent.

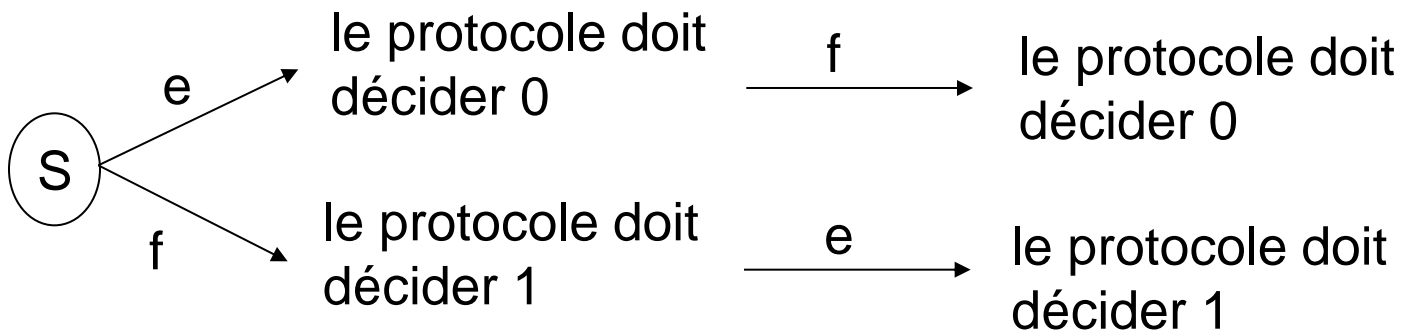


On appelle  $S$  l'état critique. On note  $e$  l'action de  $P_0$  et  $e(S)$  l'état du protocole après l'action. De même, on note  $f$  l'action de  $P_1$  et  $f(S)$  l'état du protocole après  $f$ . La valeur décidée en  $e(S)$  est différente de celle en  $f(S)$ .

# Analyse

On discute les différentes formes de e et f en se restreignant à accéder des registres atomiques en lecture ou écritures.

**Cas 1:** les actions e et f correspondent à des accès à des registres différents. Dans cette situation le résultat de e puis de f est le même que celui de f puis de e.

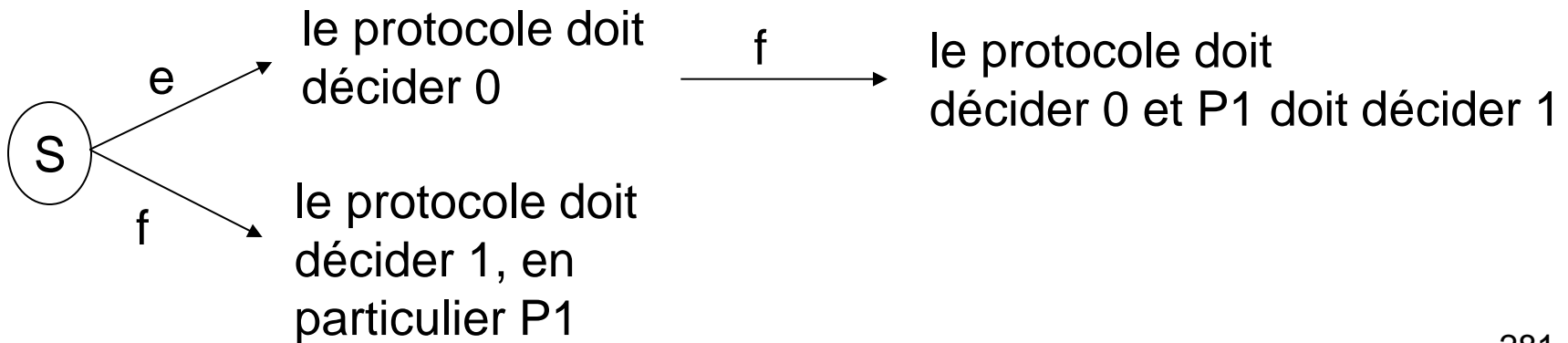




# Analyse

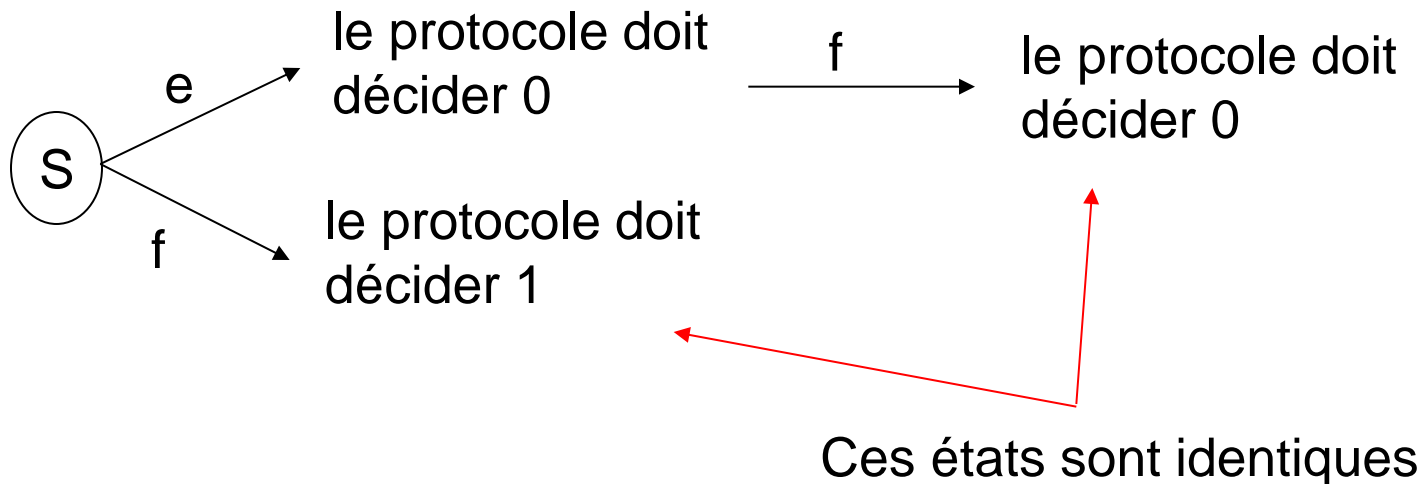
On a une contradiction car le protocole doit se trouver dans le même état après exécution de  $e$  puis  $f$  et  $f$  puis  $e$ , c'est-à-dire  $e(f(S))$  doit choisir la même valeur que  $f(e(S))$ .

**Cas 2:** Soit  $e$  soit  $f$  est une lecture, disons  $e$ . Lorsque  $P0$  exécute  $e$ , l'état de  $P1$  ne change pas. Alors l'exécution de  $P1$  depuis  $S$  ou  $e(S)$  est la même et cela contredit le fait que les valeurs choisies sont différentes



# Analyse

**Cas 3:** e et f sont des écritures dans un même registre. A nouveau les états du protocole en  $f(S)$  et  $f(e(S))$  sont identiques puisque l'écriture par P1 annule l'écriture par P0.



# Consensus – test-and-set

On rappelle l'implémentation de la fonction atomique test-and-set

```
public class TestAndSet {  
    int myValue = -1;  
  
    public synchronized int testAndSet(int newValue) {  
        int oldValue = myValue;  
        myValue = newValue;  
        return oldValue;  
    }  
}
```

Cette primitive permet de résoudre le problème du consensus pour 2 processus

# Consensus – test-and-set

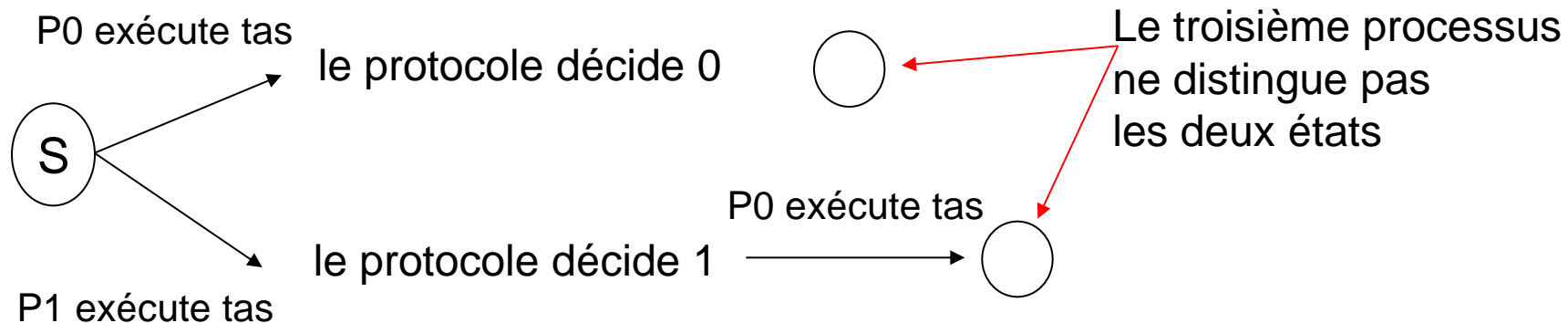
```
class TestAndSetConsensus implements Consensus {  
    TestAndSet x;  
    int proposed[] = {0, 0};  
    public void propose(int pid, int v) {  
        proposed[pid] = v;  
    }  
    public int decide(int pid) {  
        if (x.testAndSet(pid) == -1) return proposed[pid];  
        else return proposed[1-pid];  
    }  
}
```

# Consensus – test-and-set

**Corollaire:** on ne peut pas implémenter la fonction testAndSet avec des registres atomiques.

En effet, si c'était possible on pourrait résoudre le problème du consensus pour deux processus avec des registres atomiques.

On montre qu'on ne peut pas résoudre le problème du consensus pour trois processus avec la fonction testAndSet



# Consensus - Move

On considère l'opération atomique *move* qui copie la valeur d'un registre atomique dans un autre registre atomique

Cette primitive permet de résoudre le problème du consensus pour  $n$  processus avec les registres atomiques

# Consensus - Move

```
class MoveConsensus implements Consensus {  
    int proposed[] = {0, 0};  
    public void propose(int pid, int v) {  
        proposed[pid] = v; // atomique  
    }  
    public int decide(int pid) {  
        if (pid == 0) {  
            proposed[1]=proposed[0]; // atomique  
            return proposed[0];  
        }  
        else  
            move(proposed[1],proposed[0]);  
            return proposed[0];  
    }  
}
```

# Consensus - Move

Les registres `proposed[]` et l'opération `move` sont atomiques, on peut donc distinguer les situations où l'écriture dans les registres se produit avant ou après l'opération `move`.

Si `proposed[1]=proposed[0]`  $\longrightarrow$  `move(proposed[1],proposed[0])`  
alors la valeur initiale du processus 0 est choisie.

Sinon, c'est la valeur initiale du processus 1 qui est choisie.

Pour généraliser le protocole à  $n$  processus on utilise un tableau `r[0..n-1, 0..1]`. `r[i,0]=i` et `r[i,1]=i-1` initialement.



# Consensus - Move

Le processus  $i$  exécute le protocole suivant:

```
move(r[i,0],r[i,1])  
for(j=i+1; j< n; j++)  
    r[j,0] := j-1; // écriture atomique  
for(j=n-1; j>=0; j--)  
    if r[j,1]=j;  
    return j;
```

# Consensus - Move

Tous les processus décident la même valeur (agreement)

Supposons qu'un processus  $P_0$  décide la valeur  $j_0$  et un autre  $P_1$  la valeur  $j_1$  avec  $j_0 < j_1$ .

Pour décider  $j_0$ ,  $P_0$  doit lire  $r[j_1, 1] < j_1$ , c'est-à-dire que la lecture doit précéder le  $\text{move}(r[j_1, 0], r[j_1, 1])$ , de plus cette opération doit s'exécuter plus tard.

# Consensus - Move

Le numéro de  $P_0$  doit être supérieur à  $j_1$ , sinon il exécute  $r[j_1, 0] = j_1 - 1$ .

Comme il retourne  $j_0 < j_1 \leq \#P_0$ ,  $P_0$  doit nécessairement avoir  $r[\#P_0, 1] \neq \#P_0$ .

Donc un processus  $P_2$ ,  $\#P_2 < \#P_0$  a exécuté  $r[\#P_0, 0] = \#P_0 - 1$ . Avant d'exécuter cette opération, il a exécuté  $\text{move}(r[\#P_2, 0], [\#P_2, 1])$ .

Si  $\#P_2 > j_0$ , il existe un troisième processus  $P_3$ ,  $\#P_3 < \#P_2$  qui a exécuté  $r[\#P_2, 0] = \#P_2 - 1$  et ainsi de suite jusqu'à montrer qu'il existe un processus qui a exécuté  $r[j_1, 0] = j_1 - 1$ , une contradiction.

# Remarque

Le problème du consensus est utilisé pour comparer l'utilité des fonctions de synchronisation. Par exemple, on a vu qu'une fonction `testAndSet` ne peut pas être implémentée avec des registres atomiques.

Comme on peut résoudre le problème du consensus pour deux processus avec des registres atomiques, on dit qu'ils ont un numéro de consensus de 1.

La fonction `testAndSet` a un numéro de consensus de 2.

On sait qu'on ne peut pas implémenter une primitive avec un numéro de consensus  $n$  avec une primitive de numéro de consensus inférieur.

**On peut montrer l'inverse en toute généralité.** Par exemple, avec la fonction `testAndSet` on peut implémenter des registres atomiques.

# Programmation distribuée

# Modèle d'un système distribué

On considère un système distribué qui possède les caractéristiques suivantes:

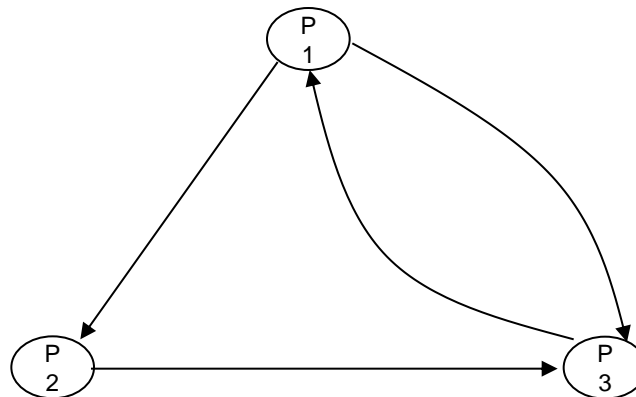
1. *Absence d'horloge commune.* Les processus s'exécutent sur des machines distantes qui possèdent des horloges locales. On ne peut pas assurer que toutes les horloges soient synchronisées.
2. *Absence de mémoire partagée.* Les processus n'ont pas accès à la totalité des données. Des algorithmes doivent être mis en œuvre pour leur permettre de connaître l'état global du système.
3. *Pas de détection de panne.* Le système que l'on considère est asynchrone. Dans ce contexte, on ne peut pas facilement faire la différence entre un processus lent et un processus en panne. Ce contexte complique les solutions des problèmes de consensus (impossibilité), élection, ....

# Modèle

Les communications se font par envois de messages asynchrones et aucune hypothèse est faite sur les délais de transmissions.

Par contre, les transmissions sont sans erreurs et les messages ne sont jamais perdus (capacité infinie).

Dans ce contexte, un programme distribué se modélise comme un graphe orienté où les sommets représentent les processus et les arcs (orientés) les canaux de transmissions.

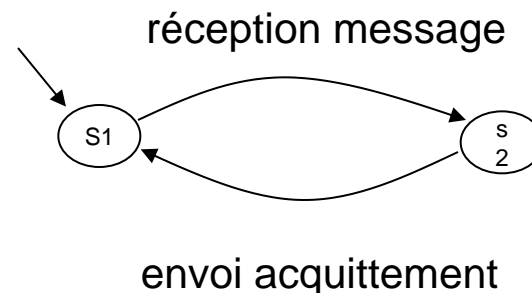
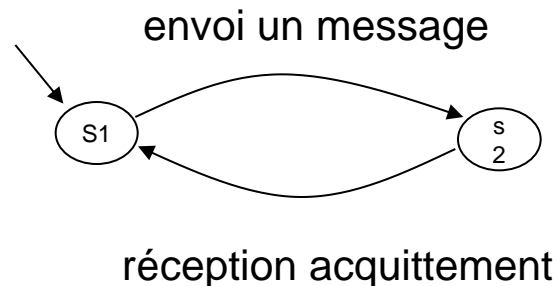


# Processus

Un processus est défini par l'ensemble de ces états, une condition initiale (un sous-ensemble de l'ensemble des états) et un ensemble d'événements.

Un événement peut changer l'état d'un processus ainsi que l'état d'un canal de transmission.

Par exemple, on modélise un système client-serveur où le serveur ne fait qu'acquitter les réceptions.





# Horloge logique

Le but est de déterminer un ordre total sur les événements en possédant une horloge qui nous permet de déterminer l'ordre d'arrivée des événements (pensez aux algorithmes rencontrés qui utilisent une estampille temporelle).

D'après la discussion autour de la relation happen-before un ordre peut-être déterminé seulement pour les instructions intra-processus et pour les instructions relatives à l'émission/réception de message.

**Une horloge logique**  $C$  est une application de l'ensemble des événements dans les entiers naturels  $C: E \rightarrow \mathbb{N}$  telle que

$$\forall e, f \in E \quad e \rightarrow f \implies C(e) < C(f)$$

# Horloge logique

On peut aussi définir une horloge logique en associant une estampille temporelle aux états des processus.

$$\forall s, t \in S \quad s \prec t \implies C(s) < C(t)$$

On impose en plus que la transmission d'un message ne prenne pas un temps nul.

**Exemple:** une horloge commune partagée par tous les processus est une horloge logique.

# Horloge logique

```
public class LamportClock {  
    int c; // l'horloge logique  
    public LamportClock() { c = 1; }  
    public int getValue() { return c }  
    public void tick() {  
        c = c + 1; // explicitement exécutée après les événements internes  
    }  
    public void sendAction() {  
        // inclure l'horloge logique dans le message  
        c = c + 1;  
    }  
    public void receiveAction(int receiveValue) {  
        c = Util.max(c, receiveValue) + 1; // après réception d'un message l'horloge  
                                           // est mise-à-jour  
    }  
}
```

# Horloge logique

On vérifie que l'algorithme satisfait

$$\forall s, t \in S \quad (s \neq t) \implies s:c < t:c$$



l'horloge c dans l'état s

Pour définir un ordre total, on inclut l'identificateur de processus pour lever les ambiguïtés possibles

$$(s:c; s:p) < (t:c; t:p) \iff (s:c < t:c) \vee ((s:c = t:c) \wedge (s:p < t:p))$$



l'identificateur du processus qui se trouve dans l'état s

# Vecteur d'horloge

L'horloge logique proposée n'assure pas que si  $s.c < t.c$  alors  $s \rightarrow t$ .  
En effet,  $(S, \rightarrow)$  définit un ordre partiel alors que l'ordre sur les entiers naturels est total.

Un vecteur d'horloge est une application  $v : S \rightarrow \mathbb{N}_k$  telle que  
 $\forall s, t \in S \quad s \rightarrow t \implies s.v < t.v$

 vecteur associé à l'état  $s$

Pour comparer les vecteurs on doit définir un ordre partiel consistant avec la relation happen-before

# Relation d'ordre

Soit deux vecteur  $x$  et  $y$  de dimension  $k$  on définit

$$x < y \iff (\exists i \in [0, k-1] \text{ tel que } x[i] < y[i] \text{ et } \forall j \in [0, i-1] \text{ tel que } x[j] = y[j])$$

$$x \leq y \iff (x < y) \vee (x = y)$$

L'ordre est partiel,  $(2, 3, 0)$  et  $(0, 4, 1)$  ne sont pas comparables.

**Une horloge vectorielle** associe à chaque état/événement un vecteur horloge.

Dans l'implémentation suivante, la taille du vecteur horloge est le nombre de processus

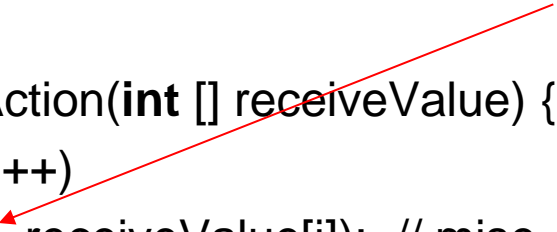
# Horloge vectorielle

```
public class VectorClock {  
    public int [] v; // vecteur horloge  
    int myId;  
    int k; // nombre de processus  
    public VectorClock( int numProc, int id) {  
        myId = id;  
        k = numProc; // une entrée dans le vecteur par horloge  
        v = new int[numProc]; // le vecteur horloge  
        for(int i = 0; i < k; i++) v[i] = 0;  
        v[myId] = 1;  
    }  
    public void tick() {  
        v[myId]++; // après chaque événement/action on incrémente l'horloge  
    }  
}
```

# Horloge vectorielle

```
public void sendAction() {  
    // inclure le vecteur dans le message  
    v[myId]++; // on incrémente son horloge  
}  
public void receiveAction(int [] receiveValue) {  
    for( int i = 0; i < k; i++)  
        v[i] = Util.max(v[i], receiveValue[i]); // mise-à-jour de toutes les entrées  
        v[myId]++; // du vecteur horloge  
}  
public int getValue(int i) { return v[i]; }  
  
public String toString() { return Util.writeArray(v); }  
}
```

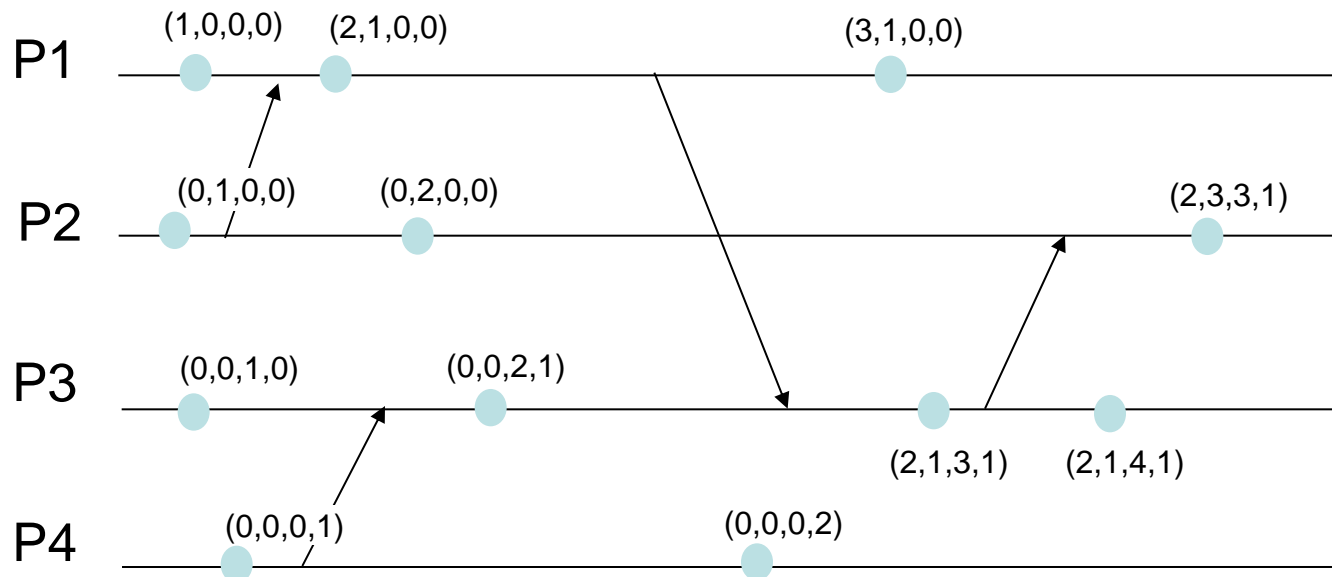
le vecteur local devient plus grand que le vecteur reçu





# Horloge vectorielle

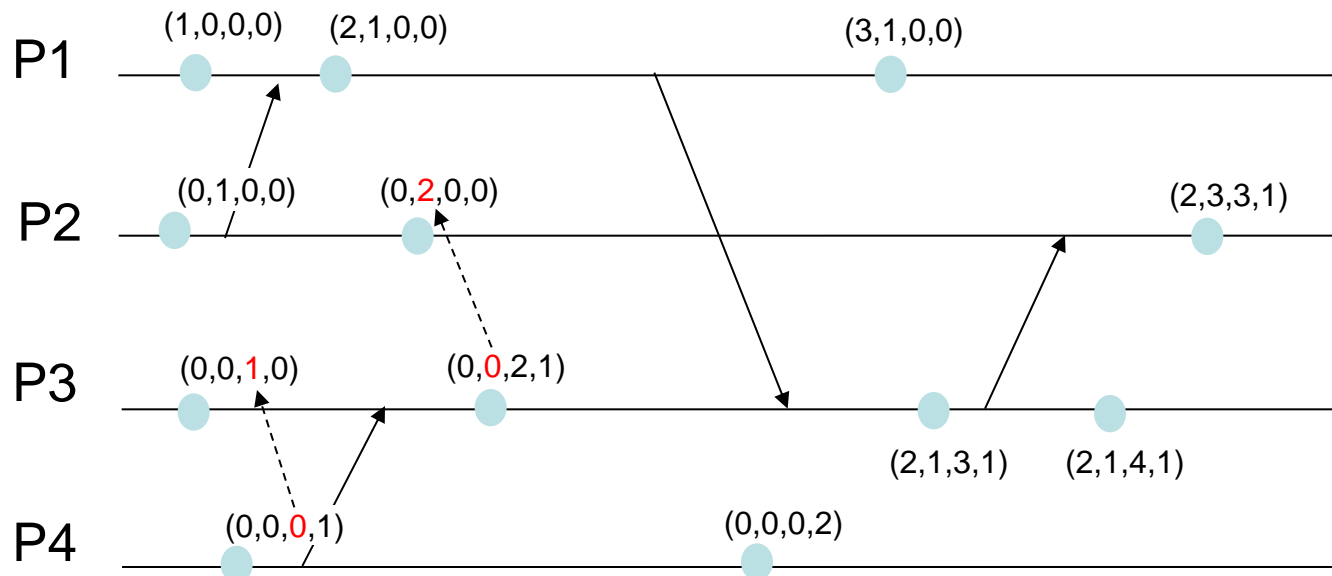
En pratique, c'est le linker qui va prendre en charge l'insertion du vecteur horloge dans le message et l'appel à la méthode *sendAction()*



# Horloge vectorielle

On montre que

si  $s \leq t$  alors  $s \leq t \Rightarrow t.v[s:p] < s.v[s:p]$



# Horloge vectorielle

**1<sup>er</sup> cas:** si  $t.p = s.p$  (même processus) alors on a nécessairement  $t \rightarrow s$   
 $t.v[t.p] < s.v[s.p]$   
 et donc

**2<sup>ème</sup> cas:**  $t.p \neq s.p$   
 $s.v[s.p]$  est l'horloge du processus  $P_{s.p}$ , comme  $s \prec t$  alors le processus  $P_{t.p}$  n'a pas connaissance de cette valeur (même pas par des processus intermédiaire sinon il y aurait une relation happen-before entre les événements). On a donc nécessairement  
 $t.v[s.p] < s.v[s.p]$

$(s \prec t) \implies (s.v < t.v)$   
 et on a montré

# Horloge vectorielle

Réciproquement, si  $s \rightarrow t$  alors il existe une séquence de processus telle que  $\mathbf{s} \rightarrow \mathbf{i} \rightarrow \mathbf{j} \dots \mathbf{k} \rightarrow \mathbf{t}$ .

$$\exists r \quad s:v[r] \cdot i:v[r] \cdot j:v[r] \cdot \dots \cdot t:v[r]$$

on a

De plus, on sait que  $t \leq s$ , on a montré que cela implique  $s:v[t:p] < t:v[t:p]$

On a donc bien

$$s \leq t \implies s:v < t:v$$

Une horloge virtuelle permet donc aux processus de déterminer la relation happen-before en cours d'exécution.

# Dépendance directe

L'horloge vectorielle nécessite d'échanger un vecteur de taille égale au nombre de processus. Pour limiter les échanges entre les processus on utilise **une horloge à dépendance directe** (*Direct Dependency Clock*). Cette horloge permet notamment de résoudre le problème de la section critique de manière distribuée.

Dans cet algorithme, les processus transmettent uniquement l'entrée de leur vecteur horloge qui correspond à leur identificateur. Cette horloge définit une relation de dépendance directe entre les processus, elle n'est pas transitive.

# Dépendance directe

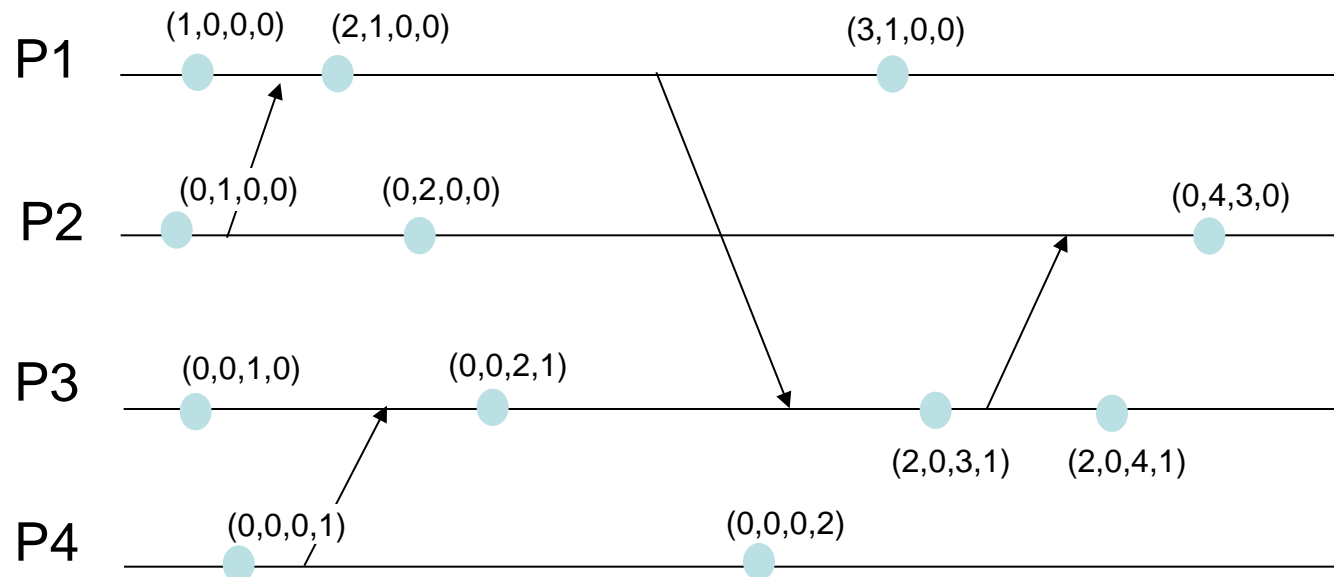
```
public class DirectClock {  
    public int [] clock;  
    int myId;  
    public DirectClock(int numProc, int id) {  
        myId = id;  
        clock = new int[numProc];  
        for( int i = 0; i < numProc; i++) clock[i] = 0;  
        clock[myId] = 1;  
    }  
    public int getValue(int i) { return clock[i]; }  
  
    public void tick() { clock[myId]++;}
```

# Dépendance directe

```
public void sendAction() {  
    // inclure clock[myId] dans le message  
    tick();  
}  
  
public void receiveAction(int sender, int receiveValue) {  
    clock[sender] = Util.max(clock[sender], receiveValue);  
    clock[myId] = Util.max(clock[myId], receiveValue) + 1;  
}  
}
```

On remarque que l'algorithme est le même que celui d'une horloge logique i on ne considère que l'entrée du vecteur qui correspond au processus qui gère l'horloge.

# Dépendance directe





# Dépendance directe

On définit une nouvelle relation partielle d'ordre entre les événements/actions que l'on note  $\rightarrow_d$  (dépendance direct).

$s \rightarrow_d t$  s'il existe un chemin allant de  $s$  à  $t$  qui emprunte au plus un message dans le diagramme happen-before.

On peut montrer que cette relation satisfait la propriété suivante

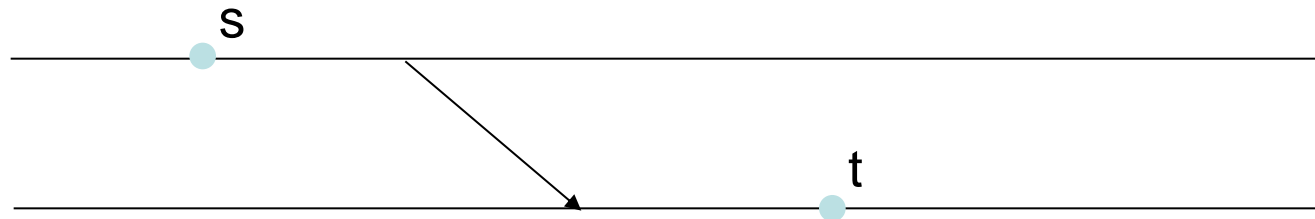
$$\exists s; t \quad s:p \in t:p \quad (s \rightarrow_d t) \iff (s:v[s:p] \cdot t:v[s:p])$$

# Dépendance directe

‘preuve’: on montre d’abord que

$$s:p \in t:p; \quad s \stackrel{!}{\rightarrow}_d t \quad ) \quad s:v[s:p] \cdot \quad t:v[s:p]$$

La relation  $\stackrel{!}{\rightarrow}_d$  indique que depuis l’état  $s$  il existe un chemin menant à l’état  $t$  qui utilise au plus un lien correspondant à l’envoi d’un message



A la réception du message, le processus  $t.p$  met à jour l’entrée de son horloge  $[s.p]$  avec une valeur supérieur à  $s.v[s.p]$ .

# Dépendance directe

$s \leq_d t$

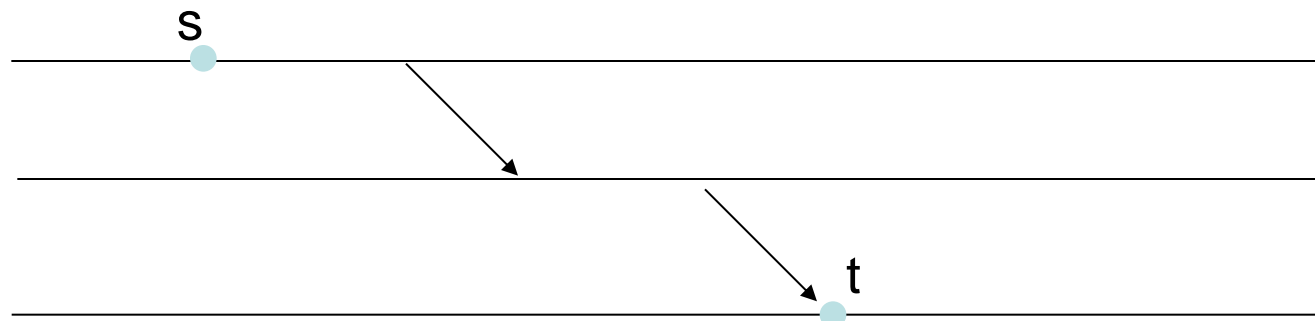
‘preuve’: On montre la réciproque. Si  $s \leq_d t$ , on a soit pas de chemin de  $s$  à  $t$  soit un chemin (minimal) qui emprunte au moins deux liens correspondant au transfert de message.

$t:v[s:p] < s:v[s:p]$

S’il n’y a pas de chemin, il est clair que l’horloge  $t:v[s:p]$  n’a pas été mise-à-jour (on a toujours  $t:v[s:p] < s:v[s:p]$  puisque  $t:v[s:p] < s:v[s:p]$  sauf pour les états qui correspondent à la transmission-réception d’un message).

$t:v[s:p]$

S’il y a un chemin indirect, le raisonnement est similaire puisque n’est pas modifié



# Application

On utilise une horloge à dépendance directe pour résoudre le problème de l'exclusion mutuelle dans un système distribué.

Chaque processus maintient une horloge logique pour estampiller les messages et une file d'attente pour mémoriser les requêtes d'accès à la section critique.

Cet algorithme (de Lamport) assure que les processus accèdent à la section critique dans l'ordre des estampilles temporelles de leurs requêtes.

# Application

- Pour accéder à la section critique un processus envoie un message avec une estampille temporelle à tous les processus et ajoute une estampille à la queue.
- Lorsqu'un processus reçoit une requête, il l'a mémorise dans la queue avec son estampille temporelle.
- Pour libérer l'accès à la section critique un processus envoie un message à tous les processus.
- Lorsqu'un processus reçoit un message de libération de la section critique, la requête correspondante est supprimée de la file d'attente.
- Un processus peut accéder à la section critique si et seulement si
  - il a déposé une requête dans la file d'attente avec une estampille  $t$
  - l'estampille  $t$  est plus petite que toutes les autres
  - il a reçu un message de tous les processus qui ont une requête dans la file d'attente (acquittement de la requête)

# Application

Pour implémenter l'algorithme on utilise deux tableaux:

- $q[j]$  qui contient l'estampille temporelle de la requête par le processus  $P_j$ .  
Si  $j$  n'a pas effectué de requête on utilise la valeur *Symbols.infinity*.

- $v[j]$  qui contient l'estampille temporelle du dernier message reçu du processus  $P_j$ . La composante  $s.v[j]$  représente la valeur de l'horloge logique dans l'état  $s$ , il s'agit d'une horloge à dépendance directe.

Plusieurs processus peuvent avoir la même estampille temporelle, on étend la relation d'ordre à relation totale en utilisant l'identificateur de processus.  $i$  entre en section critique si

$$8j; j \in I \quad (q[i]; i) < (v[j]; j) \wedge (q[i]; i) < (q[j]; j)$$

# Algorithme de Ricart et Agrawala

L'algorithme précédent assure l'exclusion mutuelle, utilise une horloge à dépendance directe et nécessite l'envoi de  $3(N-1)$  messages ( $N-1$  pour la requête,  $N-1$  acquittement et  $N-1$  libération).

L'algorithme de Ricart et Agrawala nécessite l'envoi de  $2(N-1)$  messages. Il utilise un horloge de Lamport c'est-à-dire un entier qui est mis-à-jour à chaque réception de message.

# Algorithme de Ricart et Agrawala

- Pour faire une requête un processus transmet un message estampillé *request*.
- Après réception d'une requête un processus transmet un message *okay* si il n'est pas intéressé par accéder la section critique ou si l'estampille de sa requête à une valeur plus grande. Sinon, l'identificateur du processus est mémorisé et le message *okay* sera transmis plus tard.
- Pour libérer la section critique un processus transmet un message *okay* à tous les processus.
- Un processus peut accéder la section critique si et seulement si il a effectué une requête et reçu un message d'acquiescement de la part de tous les autres processus



# Algorithme de Ricart et Agrawala

```
int myts; // estampille temporelle locale
LamportClock c = new LamportClock(); // horloge logique
IntLinkedList pendingQ = new IntLinkedList(); // liste de proc. en attente
int numOkay; // nombre de message d'acquittement reçus

public RAMutex(...) {
    ...
    myts = Symbols.Infinity; // pas d'intérêt à accéder la section critique
}

public synchronized void requestCS() { // des processus indépendants invoquent
    requestCS et handleMsg
    c.tick(); myts = c.getValue(); broadcastMsg(' request ', myts); numOkay = 0;
    while (numOkay < N-1) myWait();
}
```

# Algorithme de Ricart et Agrawala

```
public synchronized void releaseCS() {  
    myts = Symbols.Infinity;  
    while (! pendingQ.isEmpty()) {  
        int pid = pendingQ.removeHead();  
        sendMsg(pid, ' okay ', c.getValue());  
    }  
}  
  
public synchronized void handleMsg(Msg m, int src, String tag) {  
    int timeStamp = m.getMessageInt(); // récupération de l'estampille du message  
    c.receiveAction(timeStamp); // mis-à-jour de l'horloge max(c, timeStamp) + 1  
    if (tag.equals(' request')) {  
        if ((myts == Symbols.Infinity) || (timeStamp < myts) ||  
            ((timeStamp == myts) && (src < myId)))  
            sendMsg(src, ' okay ', c.getValue());  
        else pendingQ.add(src);  
    } else if (tag.equals(' okay ')) notify(); // voir requestCS()  
}
```

# Algorithme de Ricart et Agrawala

On montre que cet algorithme assure l'exclusion mutuelle.

Supposons que les processus  $i$  et  $j$  se trouvent en section critique simultanément.

Ces processus ont déterminé une valeur  $myts\_i$  et  $myts\_j$  avant d'accéder à la section critique et envoyer cette valeur à tous les processus.

**Cas 1:** le processus  $j$  choisit  $myts\_j$  *après* avoir répondu au processus  $i$  (méthode `handleMsg` qui est synchronisée).

$i$  choisit  $myts\_i \rightarrow i$  envoie la requête  $\rightarrow j$  exécute `handleMsg`  $\rightarrow j$  répond au processus  $i$  (*okay*)  $\rightarrow j$  choisit  $myts\_j$

Avant de répondre le processus  $j$  met à jour son horloge et donc  $myts\_j > myts\_i$ . Le processus  $i$  ne répondra pas à la demande du processus  $j$  avant de sortir de la section critique et les deux processus ne peuvent pas s'y trouver simultanément.

# Algorithme de Ricart et Agrawala

**Cas 2:** Le processus i choisit  $myts\_i$  après avoir répondu au processus j. Ce cas est similaire au premier.

**Cas 3:** Les processus i et j choisissent  $myts\_i$  et  $myts\_j$  avant de se répondre.

requestCS\_i choisit  $myts\_i$  → envoi la requête  
hMsg\_i

requestCS\_j choisit  $myts\_j$  → envoi la requête  
hMsg\_j

réception → réponse

réception → réponse

Les deux processus ne peuvent pas répondre les deux. En effet, les processus i et j répondent si  $myts\_i << myts\_j$  et  $myts\_j << myts\_i$ .

# Algorithme de Ricart et Agrawala

L'algorithme assure qu'un processus qui exécute `requestCS` entrera en section critique (pas d'insuffisance de ressource).

Le processus exécute `requestCS_i`, il détermine `myts_i` et envoie une requête à tous les processus. A un instant  $t$ , toutes les requêtes seront reçues et les processus mettront à jour leur horloge (`c.receiveAction(...)`).

Soit  $V$  l'ensemble des processus qui ont demandé à entrer en section critique avec une estampille temporelle inférieure ou égale à `myts_i`. Le cardinal de cet ensemble ne peut pas augmenter après le temps  $t$ .

L'ensemble des estampilles temporelles associés au processus de l'ensemble  $V$  sont ordonnées (ordre  $\ll$  total). La requête associée au processus possédant l'estampille la plus petite va être traitée par tous les processus qui vont y répondre par un message *okay*. Le cardinal  $V$  va donc diminuer. Par récurrence on montre que tous les processus vont accéder à la section critique dans l'ordre de leur estampille temporelle.

# Autres stratégies pour l'exclusion mutuelle

Une stratégie classique pour résoudre le problème de l'exclusion mutuelle dans les systèmes distribués consiste à utiliser un jeton (token). Seul le processus qui dispose du jeton est habilité à accéder la section critique.

On peut implémenter cette idée en utilisant un algorithme **centralisé**.

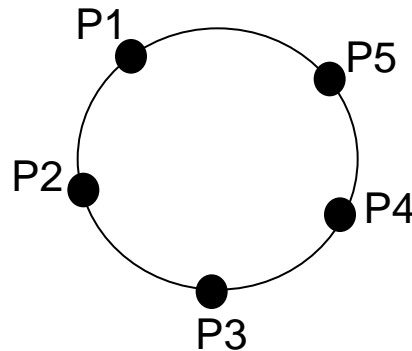
Un processus est désigné (comment?) pour être le coordinateur.

Chaque processus qui désire accéder la section critique envoie un message au coordinateur qui place le message dans une file FIFO.

Lorsqu'un processus sort de la section critique il informe le coordinateur qui envoie un message au processus prioritaire dans le file FIFO.

# Autres stratégies pour l'exclusion mutuelle

Une autre implémentation consiste à définir un ordre cyclique sur les processus, la topologie devient celle d'un *anneau*.



Un message particulier, *le jeton*, circule de processus en processus. Seul le processus qui dispose du jeton peut accéder la section critique.

Le jeton peut aussi se déplacer seulement à la réception d'une requête.

# Quorum

Les algorithmes qui utilisent un jeton sont vulnérables aux pannes du processus qui dispose du jeton.

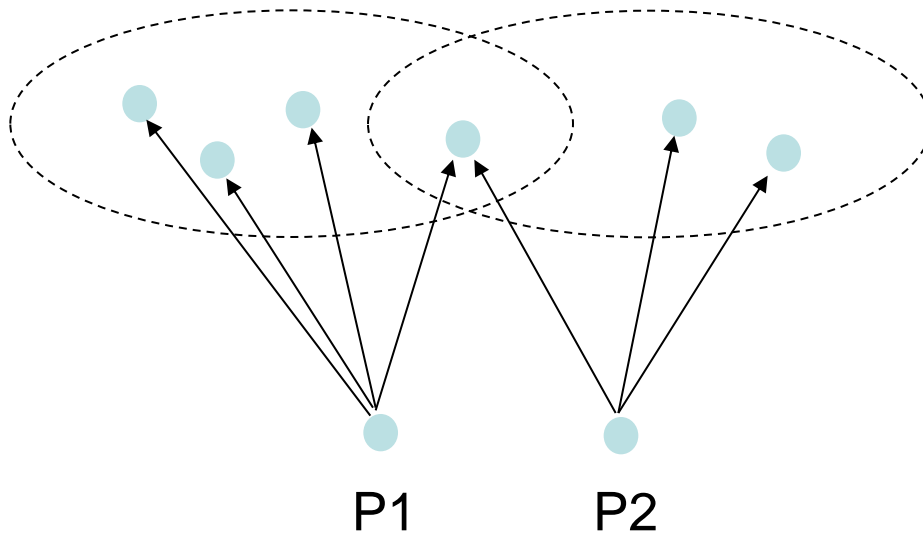
Les algorithmes qui utilise une estampille temporelles génèrent beaucoup de messages.

Une solution intermédiaire consiste à associer à chaque processus un sous-ensemble de processus, *request set*, auxquels le processus va demander la permission d'accéder la section critique.

Si ces sous-ensembles de processus ont des intersections non vides, on assure l'exclusion mutuelle. Typiquement on utilise des ensembles de  $\lceil (N+1)/2 \rceil$  processus.



# Quorum



# Problème du consensus

Le problème du consensus dans les systèmes distribués s'énonce comme dans le cas de systèmes à mémoire partagée. Chaque processus possède une valeur initiale. Après exécution de l'algorithme tous les processus décident d'une valeur telle que

1. Tous les processus actifs (non en panne) décident de la même valeur (agreement).
2. Toutes les valeurs initiales peuvent être choisies (pour éviter la solution triviale) (validity)
3. Les processus actifs doivent décider d'une valeur en un temps fini (termination).

Dans l'énoncé des caractéristiques du protocole pour le problème du consensus, on distingue les processus actifs des processus qui peuvent être 'en panne'. On distingue plusieurs types de pannes.

# Types de pannes

1. **Crash**, Un processus en panne arrête de s'exécuter et n'effectuera plus d'action dans le futur. Ce type de panne n'est pas détectable dans les systèmes asynchrones.
2. **Crash + lien**, soit un processus arrête de s'exécuter soit un lien de communication devient inutilisable. On distingue deux sous-cas, dans le premier les processus restent connectés dans le second ils deviennent déconnectés.
3. **Omission**, un processus omet de transmettre ou de recevoir un sous-ensemble de message qu'il aurait du recevoir/transmettre en exécutant le protocole.
4. **Byzantine**, un processus à un comportement imprévisible, des messages peuvent être transmis avec des valeurs fausses, différents messages peuvent être transmis à différents processus, ...

# Systèmes synchrones

Un résultat de Fischer, Lynch et Paterson (FLP) montre que le **problème du consensus qui tolère la panne (crash) d'un processus ne peut pas être résolu dans un système asynchrone**. La démonstration de ce résultat est 'assez' similaire à celle que l'on a vu dans les systèmes à mémoire partagée.

Par contre, on peut résoudre ce problème dans les systèmes synchrones si on peut borner le nombre de processus susceptibles de tomber en panne.

**Rappel:** Un système est synchrone si on peut borner le temps qui sépare l'émission d'un message de sa réception.

# Consensus

Chaque processus exécute le protocole suivant

$P_i$  ::

**var**

$V$  : ensemble des valeurs reçues, initialement  $\{v_i\}$

**for**  $k:=1$  **to**  $f+1$  **do**

transmettre à tous les processus

$V \cup \{v_i\}$  telles que  $P_i$  n'a pas encore transmis  $v_i$

recevoir  $S_j$  de tous les processus  $P_j; j \in I$

**end**

$y = \min(V)$

choisir

# Consensus

Comme le système est supposé synchrone, la réception de tous les messages utilise un temporisateur. Après avoir envoyé les messages, il attend un temps fini la réponse des autres processus.

Le protocole se termine donc en un temps fini (termination).

le protocole est valide car la valeur choisie appartient à l'ensemble  $V$ .

Le protocole supporte au plus la panne de  $f$  processus.

En effet, soit  $V_i$  l'ensemble des valeurs reçues par le processus  $P_i$  après les  $f+1$  transmissions/réceptions.

On montre que si  $x \in V_i$  alors  $x \in V_j$  pour  $P_j$  un processus correct, c'est-à-dire qui a exécuté le protocole jusqu'à la fin.

# Consensus

**Cas 1:** on suppose que la valeur  $x$  à été ajoutée à l'ensemble  $V_i$  au tour  $k < f + 1$ . Dans cette situation, le processus  $i$  a transmis la valeur  $x$  à tous les processus corrects.

$$k = f + 1$$

**Cas 2:** la valeur  $x$  à été ajoutée au tour  $k$

Il existe donc un processus qui a reçu cette valeur au tour  $f$  et transmise

Il existe donc un processus qui a reçu cette valeur au tour  $f-1$  et transmise

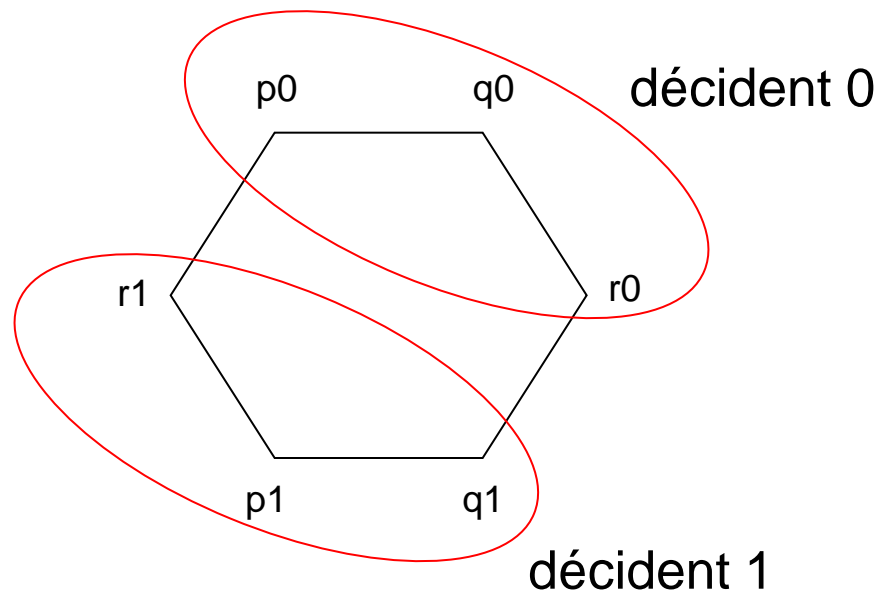
...

Il existe donc  $f+1$  processus distincts qui ont reçus cette valeur et l'ont transmise. Par hypothèse le nombre de processus qui tombe en panne est plus petit que  $f+1$ , il y a donc un processus correct qui a reçu cette valeur et l'a transmise à tous les processus (corrects).

# Erreur Byzantine

On considère le problème du consensus en tolérant des erreurs de type Byzantine de la part des processus.

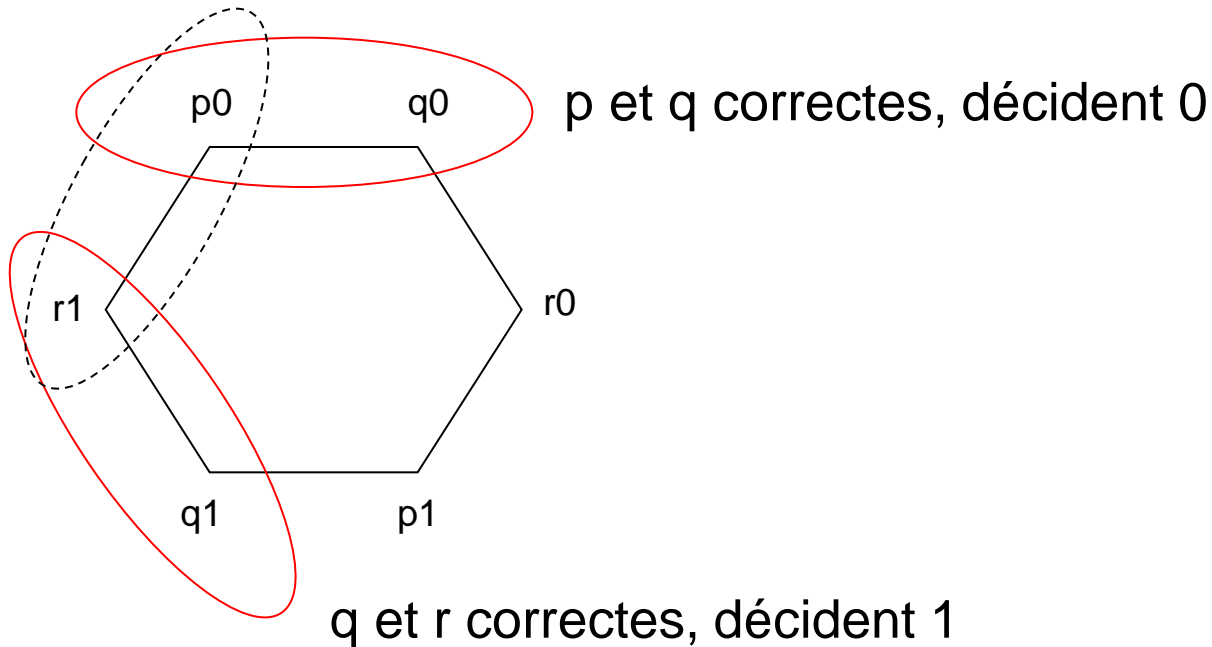
Avec trois processus, le problème ne peut pas tolérer de panne.





# Erreur Byzantine

quelle valeur choisir? dépend de q qui peut transmettre 0 à p et 1 à r



De manière générale, on ne peut pas résoudre ce problème si  $n \leq 3f$

# Erreur Byzantine

On considère le problème du consensus comme un problème de décision. A un ensemble de valeurs correspondant aux valeurs annoncées par les processus, on doit associer une valeur.

$$0\ 0\ 0 \rightarrow 0$$

$$1\ 1\ 1 \rightarrow 1$$

$$0\ 0\ * \rightarrow 0 \text{ pour tolérer une défaillance de } r$$

$$*\ 1\ 1 \rightarrow 1 \text{ pour tolérer une défaillance de } p$$

Quelle valeur associer à  $0\ * \ 1$ ?

$0\ * \ 1 \rightarrow 0$  et la valeur reçue par  $q$  est 1 ce n'est pas cohérent avec  $0\ 1\ 1 \rightarrow 1$

$0\ * \ 1 \rightarrow 1$  et la valeur reçue par  $q$  est 0 ce n'est pas cohérent avec  $0\ 0\ 1 \rightarrow 0$

# Coordinateur tournant

On peut résoudre ce problème si on suppose que le nombre de processus défaillant  $f$  satisfait  $N > 4f$ .

L'algorithme procède à  $f+1$  étapes.

A chaque étape un processus est désigné comme le coordinateur.

Chaque processus mémorise une valeur de préférence qui est initialement sa valeur initiale.

Pendant une étape, les processus échangent leur valeur de préférence. Chaque processus choisit comme valeur de préférence la valeur qu'il a reçue le plus de fois -> mise à jour de *myValue*

Ensuite, chaque processus reçoit la (nouvelle) valeur du coordinateur *kingValue*.

# Coordinateur tournant

Les processus peuvent ne pas recevoir la valeur du coordinateur si ce dernier est défaillant. Dans ce cas, ils choisissent une valeur par défaut.

Si le nombre de valeur reçues qui déterminent  $myValue \leq N/2 + f$  alors le processus exécute  $myValue = kingValue$ . Sinon, la valeur n'est pas modifiée.

# Code partiel

```
public synchronized void handleMsg(Msg m, int src, String tag) {  
    if (tag.equals(« phase 1 »)) V[src] = m.getMessageInt();  
    else if (tag.equals(« king »)) kingValue = m.getMessageInt();  
}
```

// exécutée en premier pour le consensus

```
public synchronized void propose (int val) {  
    for(int i = 0; i < N; i++) V[i] = defaultValue;  
    V[myId] = val;  
}
```

# Code partiel

```
public int decide() {  
    for(int k = 0; k<=f; k++) { // f+1 étapes  
        broadcastMsg(« phase 1 », V[myId]);  
        Util.mySleep(Symbols.roundTime); // réseau synchrone  
        synchronized(this) {  
            myValue = getMajority(V);  
            if (k == myId) broadcast(« king », myValue);  
        }  
        Util.mySleep(Symbols.roundTime);  
        synchronized(this) {  
            if (numCopies(V, myValue) > N/2 + f) V[myId] = myValue;  
            else V[myId] = kingValue;  
        }  
    }  
}
```

# Analyse de l'algorithme

Supposons qu'il existe une étape pendant laquelle tous les processus corrects choisissent la même valeur. Cette valeur persiste jusqu'à la fin de l'exécution du protocole. En effet,

$$N > 4f, \quad N \geq N/2 > 2f, \quad N \geq f > N/2 + f$$

Le nombre de valeur échangée par le processus corrects est suffisamment grand pour que la valeur *kingValue* ne soit jamais considérée.

Comme le nombre d'étape est  $f+1$ , pendant au moins une étape le coordinateur est un processus correct. Durant l'étape correspondante un processus choisit *myValue* si et seulement si il a reçu au moins  $N/2 + f + 1$  telle valeur et le coordinateur choisit aussi cette valeur (au moins  $N/2 + 1$  réception) ainsi que tous les processus correctes.

# Transactions

**Une transaction** est composée d'un ensemble d'actions (d'opérations) telles que la séquence apparait comme une action indivisible.

Pour un observateur, la transaction apparait comme exécutée ou non (l'ensembles de opérations sont exécutées ou aucune opération est exécutée).

Plusieurs transactions peuvent s'effectuer de manière concurrentes et indivisibles.

Lorsqu'une transaction est réalisée, le résultat de ses opérations est permanent.



# Exemple

On considère une transaction qui consiste à transférer une somme d'argent d'un compte A vers un compte B.

La transaction se décompose:

1. débiter le compte A
2. transférer l'argent vers le gestionnaire du compte B (message).
3. créditer le compte B

compte A

800

600

compte B

600

800

Une transaction T2 qui retourne la somme des comptes A et B retourne toujours 1400.

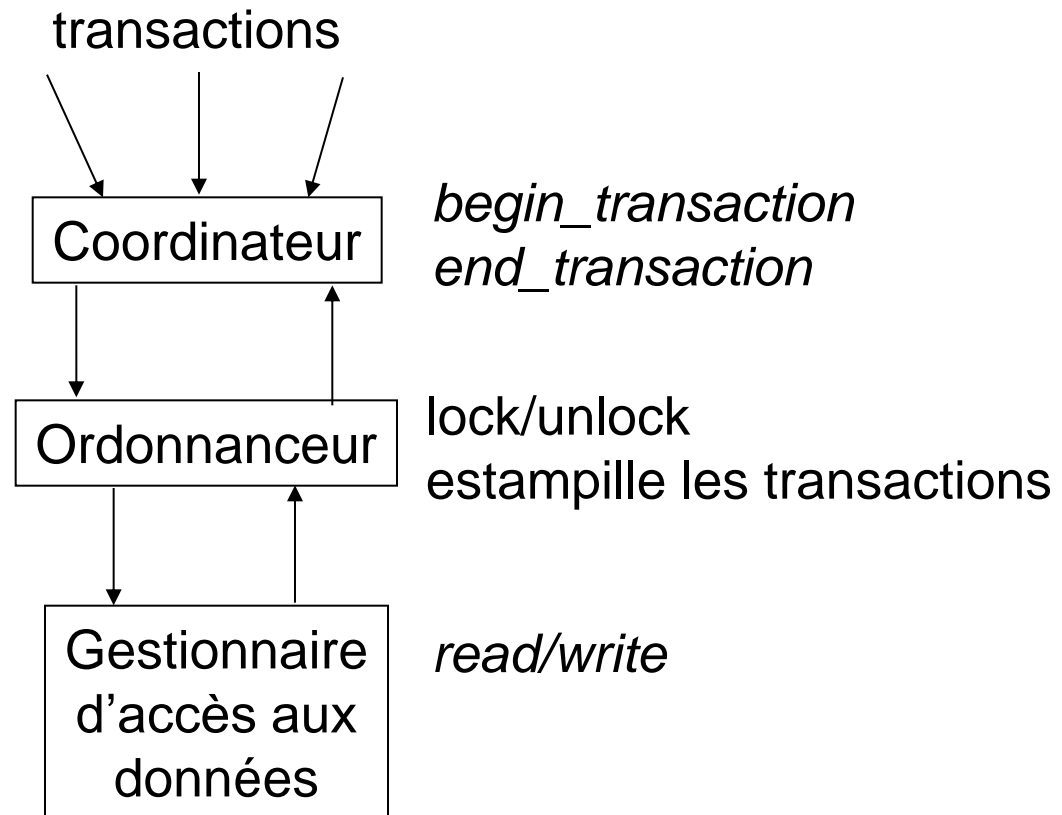
# Exemple

Si une erreur survient entre le moment où le compte A a été débité et le moment où le compte B a été débité la transaction n'est pas *confirmée* (committed), le compte A n'est pas modifié.

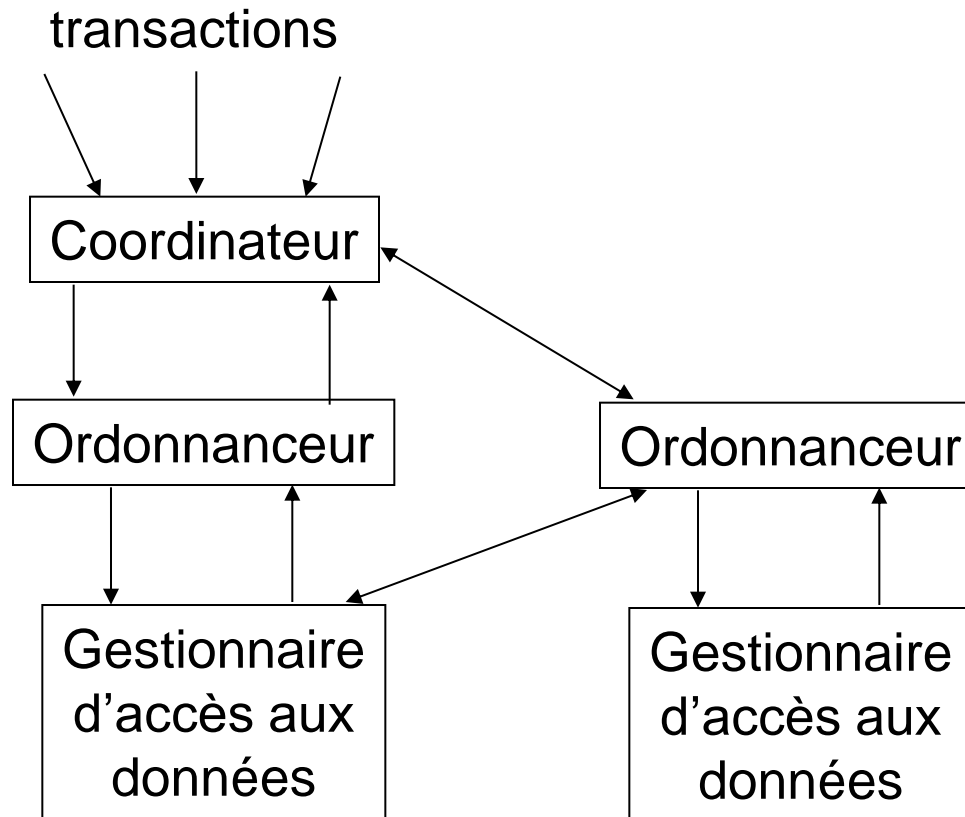
Une transaction (sur un objet) doit implémenter les méthodes suivantes:

1. *begin\_transaction*
2. *end\_transaction*
3. *abort\_transaction*
4. *read*
5. *write*

# Organisation



# Organisation



# Propriétés

Les propriétés qui doivent être assurées par une transaction sont:

1. **Atomicity**: la transaction apparaît comme une opération atomique. D'autres processus ne peuvent pas 'voir' les états intermédiaires.
2. **Consistency**: Une transaction doit respecter les contraintes d'intégrité du système. Par exemple, le transfert d'argent entre deux compte doit préserver la somme totale d'argent dans le système. La consistance est définie par des invariants.
3. **Isolation**: une transaction est isolée de l'effet d'une autre transaction, il n'y a pas d'interférence entre les transactions. L'effet de plusieurs transactions concurrentes est identique à l'exécution des transaction séquentiellement (sérialisable)
4. **Durability**: Une fois qu'une transaction est confirmée son effet est permanent.

On se réfère à ces propriétés en utilisant l'acronyme ACID.

# Exemple

On considère trois compte A, B et C, chaque compte est crédité de 1000.-. On désire créditer 100.- de A vers B, 200.- de B vers C. Le solde final des trois compte doit être de 900.-, 900.- et 1200.-

Transaction A  
*begin\_transaction*  
*x = read(A)* 1  
*x = x-100* 2  
*write(x,A)* 3  
*x = read(B)* 4  
*x = x + 100* 11  
*write(x,B)* 12  
*end\_transaction*

Transaction B  
*begin\_transaction*  
*y = read(B)* 5  
*y = y-200* 6  
*write(y,B)* 7  
*y = read(C)* 8  
*y = y + 200* 9  
*write(y,C)* 10  
*end\_transaction*

*L'exécution en rouge donne les soldes 900.-, 1100.-, 1200.-, qui est incorrecte*

# Exemple

Pour assurer l'atomicité on peut avant d'exécuter une transaction bloqué l'accès à la base de donnée. Ainsi, toutes les exécutions sont réellement séquentielles.

Une meilleure technique consiste à décomposer la transaction en deux phases. Une première pendant laquelle on bloque l'accès aux comptes et une deuxième phase pendant laquelle on débloque les accès.

Les verrous peuvent être associé à une donnée particulière, plus généralement on associe un verrou à toutes les données qui satisfont un prédicat logique (valeur inférieur à ...)

# Exemple – two-phase locking

Transaction A  
*begin\_transaction*  
*lock(A)*  
*x = read(A)*  
*x = x - 100*  
*write(x,A)*  
*lock(B)*  
*x = read(B)*  
*x = x + 100*  
*write(x,B)*  
*unlock(A)*  
*unlock(B)*  
*end\_transaction*



# Ordonnanceur – two-phase locking

L'ordonnanceur reçoit des requêtes correspondant à des opérations à effectuer dans la base de donnée  $oper(T,x)$ , T est la transaction et x la donnée.

1. Si l'opération est en conflit avec une autre opération la requête est temporisée. Sinon, un verrou sur la donnée x est donné à la transaction T et l'opération est transmise au gestionnaire d'accès des données
2. L'ordonnanceur libère le verrou seulement lorsque le gestionnaire l'a informé que l'opération est terminée
3. Une fois que l'ordonnanceur a libéré un verrou associé à une transaction il ne va plus jamais donner de verrou à cette transaction.

Attribuer les verrous selon cette politique assure que les données accédées sont toujours cohérentes.

# Ordonnanceur – estampilles temporelles

Une approche différente consiste à utiliser une horloge logique. A chaque transaction est associé une *estampille temporelle*  $ts(T)$  et chaque opération de  $T$  est estampillée avec  $ts(T)$ .

A chaque donnée sont associés deux estampilles temporelles  $tsRD(x)$  qui est la date de la dernière lecture et  $tsWR(x)$  qui est la date de la dernière écriture.

Lorsque l'ordonnanceur considère une opération  $read(T, x)$ , il l'exécute seulement si  $ts(T) > tsWR(x)$  et exécute  $tsRD(x) = \max(ts(T), tsRD(x))$ .

Si  $ts(T) < tsWR(x)$  une transaction qui a commencé après  $T$  a modifié  $x$  et la transaction  $T$  est annulée (*abort*). L

# Ordonnanceur – estampilles temporelles

De même, si l'ordonnanceur considère une opération  $write(T, x)$ , il l'exécute seulement si  $ts(T) > tsRD(X)$ . Si  $ts(T) < tsRD(X)$  la transaction  $T$  est annulée (*abort*) car la donnée  $x$  a été lue par une transaction plus récente que  $T$ .

En pratique, les estampilles temporelles permettent de limiter le temps de gestion des verrous. De plus, cette politique ne peut pas conduire à des interblocages.

# Récupérer les erreurs

Pour récupérer l'état de la base de données après d'éventuels erreurs on considère deux techniques courantes:

**Espace de travail privé** (private workspace): Une transaction n'affecte pas les données originales. Les objets modifiés pas la transaction sont copiés.

Si la transaction est interrompue (abort) ces copies sont détruites.

Si la transaction est confirmée, ces copies deviennent les originaux.

**journal des opérations** (logging): Les mise-à-jour sont effectuées directement dans la base de données. Une journal des écritures est maintenu. Si la transaction doit être interrompue on peut remettre la base de données dans l'état original.

# Confirmation d'une transaction

Lorsque plusieurs sites sont impliqués dans une même transaction ils doivent tous décider d'annuler la transaction ou de la confirmer.

L'algorithme utilisé doit être tolérant aux pannes et satisfaire:

1. Tous les processus corrects prennent la même décision
2. Si un processus décide d'annuler la transaction, tous les processus doivent l'annuler.
3. S'il n'y a pas de pannes, tous les processus doivent prendre une décision.
4. Les processus corrects doivent prendre une décision.

On considère l'algorithme suivant qui suppose que les liens de communication sont fiables et satisfait les trois premières contraintes.

# Confirmation distribuée

1. Le coordinateur envoie un message *request* à tous les sites
2. Chaque site répond avec un message *yes* pour signifier au coordinateur qu'il peut confirmer la transaction, *no* dans le cas contraire
3. Le coordinateur attend les réponses de tous les sites, s'il reçoit au moins un message *no* il envoie un message *finalAbort*, sinon il confirme la transaction en envoyant un message *finalCommit*.
4. Les sites exécutent l'action correspondante.

Un site peut tomber en panne après avoir transmis un message *yes*. Dans cette situation, après remise en route (le canal de transmission est fiable) il doit être capable de confirmer la transaction.

# Confirmation distribuée

Si un site ne reçoit pas de message *request* (il utilise un temporisateur) il annule la transaction et transmet un message *no* au coordinateur.

Le coordinateur peut aussi tomber en panne après avoir transmis un message *request*. Dans cette situation, un site qui a transmis un message *yes* doit contacter tous les sites impliqués dans la transactions et ils doivent se mettre d'accord (consensus) si oui ou non la transaction peut être confirmée. Si au moins un site a reçu un message *no*, la transaction doit être annulée.

# Instantané global

Certaines applications requièrent d'obtenir une vision globale à un instant donné de l'état d'un système distribué. Par exemple, pour réaliser un inventaire.

Pour connaître l'état global courant du système, le système doit cesser de fonctionner.

Une autre stratégie consiste à obtenir un état global du système à un instant dans le *passé*. Par exemple, pour déterminer des propriétés qui sont stables (une fois réalisées elles le sont de manière permanente) il est suffisant d'obtenir des informations du système à des instants passés (déterminer s'il y a interblocage, perte du jeton, ...)

Un tel algorithme réalise un instantané global du système (*global snapshot*).



# Instantanée global

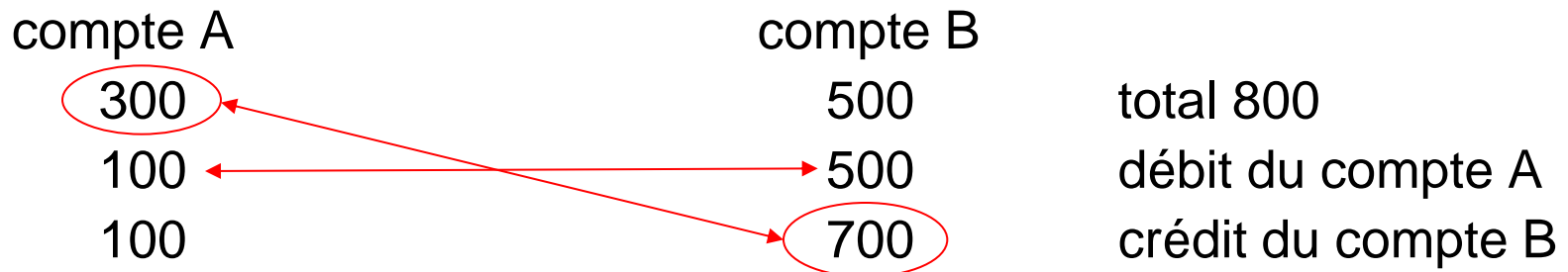
Une première difficulté est de définir ce qu'est un état global. Une première approche consiste à le définir comme étant un ensemble d'états locaux (les processus) à un instant donné. Cette définition suppose que l'on est capable de définir un même instant pour tous les processus (horloge globale).

Dans le modèle *happen-before*, un état global est un ensemble d'états locaux tous concurrents. Cette définition nécessite aucune hypothèse particulière sur le système pour pouvoir être utilisée (synchronisation, horloge globale).

# Exemple

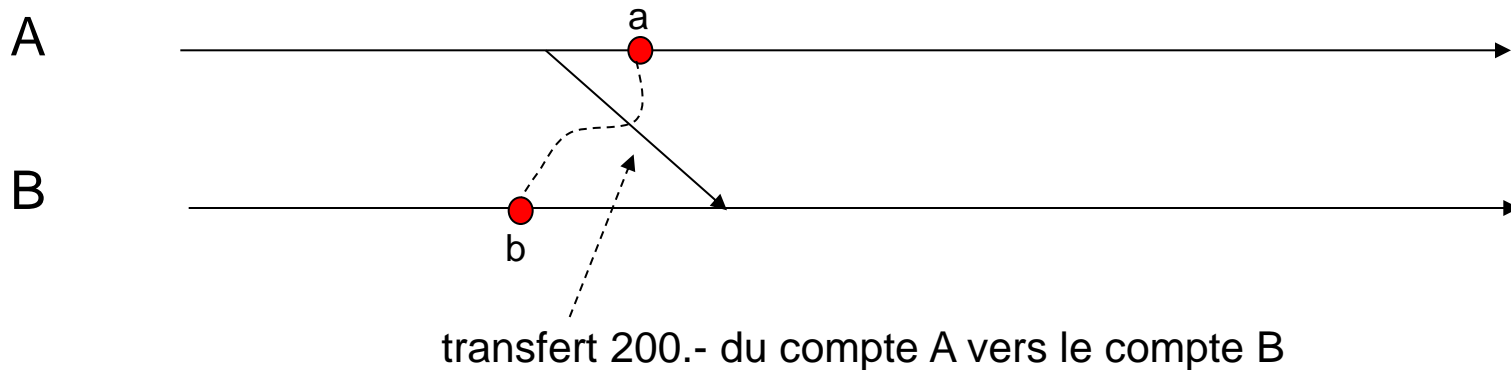
On considère un client qui possède 2 comptes en banques avec 300.- et 500.- sur chacun. En l'absence de mouvement, un instantané global doit nous permettre de conclure que le client possède au total 800.-.

Que ce passe-t-il si un transfert de 200.- est exécuté



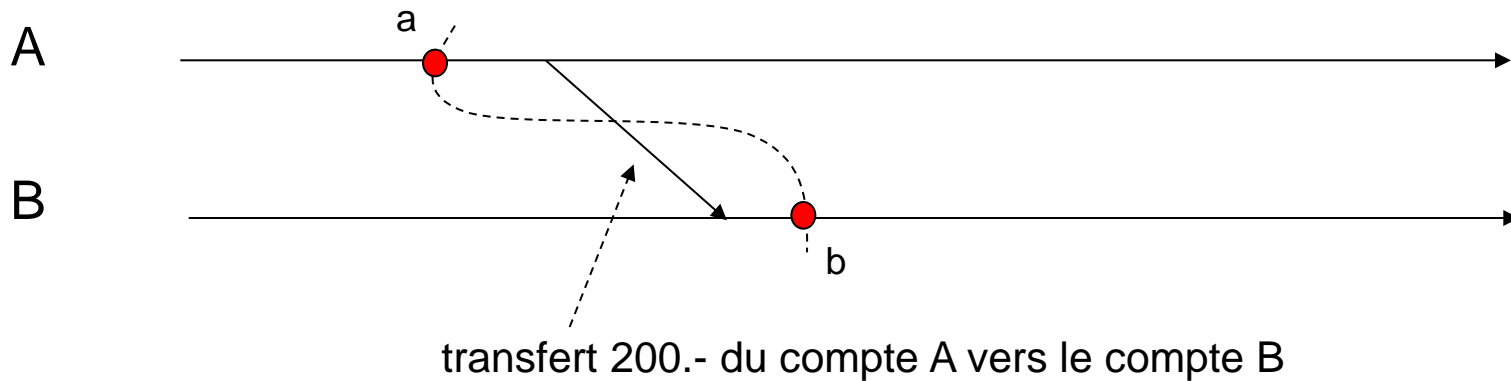
L'instantané global doit être cohérent avec la relation happen-before, de plus **les données transmises doivent être incluses dans l'instantané global.**

# Exemple



Un exemple d'état global consistant. On connaît le solde du compte A (100.-), on mémorise qu'un message à été transmis correspondant au transfert de 200.- et on connaît le solde du compte B avant d'être crédité (500.-).

# Exemple



La donnée des deux états a et b ainsi que le message correspondant au transfert n'est pas cohérent. Dans l'état a, le message n'est pas envoyé (le compte A pas nécessairement débité).

# Définition

On suppose donné une exécution  $(E, \rightarrow)$ , c'est-à-dire un ensemble d'événements et une relation de précédence. Les événements d'un même processus sont ordonnés avec une relation totale  $<$ .

On définit une coupe comme étant un sous-ensemble  $F \subseteq E$  satisfaisant

$$f \in F \wedge e < f \Rightarrow e \in F$$

Une coupe consistante ou un instantané global est un sous-ensemble  $F \subseteq E$  satisfaisant

$$f \in F \wedge e \neq f \Rightarrow e \in F$$

# Algorithme de Chandy et Lamport

On suppose que les canaux de communications sont unidirectionnels et FIFO.

Les interfaces à implémenter sont:

```
public interface Camera extends MsgHandler {  
    void globalState(); // pour initier l'intsantané global  
}
```

```
public interface CamUser extends MsgHandler {  
    void localState(); // mémorise l'état d'un processus  
}
```

# Algorithme de Chandy et Lamport

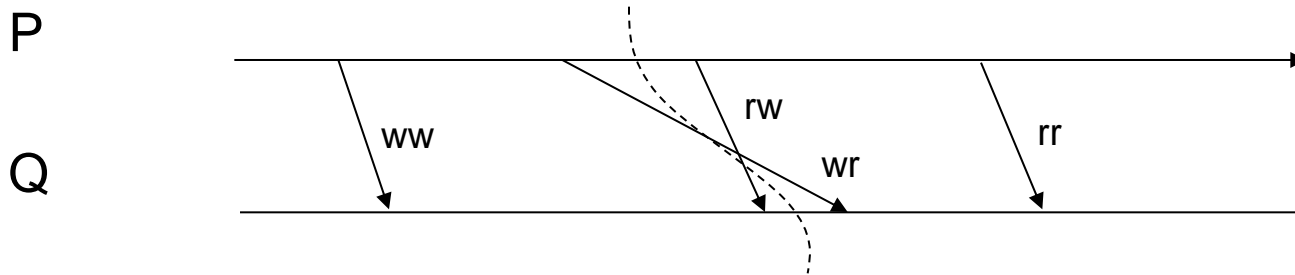
L'idée de l'algorithme est d'associer à chaque processus deux couleurs, blanc et rouge. L'état d'un processus associé à l'instantané global est celui dans lequel se trouve le processus juste avant de passer de blanc à rouge.

Un processus qui initie l'algorithme transmet des requêtes aux autres processus qui mémorisent leur état local et change leur couleur (blanc → rouge). Les changements de couleurs doivent vérifier

1. Les états locaux doivent être concurrents
2. L'état du canal doit être connu.

Pour cela, les processus utilisent un message dédié (*marker*), qui est transmis à tous les autres processus *avant* de transmettre les messages courants. Un processus qui reçoit un message *marker* devient rouge (ou le reste). On s'assure ainsi qu'un processus blanc ne reçoit jamais de message d'un processus rouge

# Algorithme de Chandy et Lamport



messages ww: messages échangés avant que les processus soient impliqués

messages rr: messages échangés après l'instantané global

messages rw: ces messages correspondent à un instantané inconsistent, le *marker* prévient l'existence de tels messages

messages wr: ces messages composent l'état global du système.

Un processus  $P_j$  mémorise tous les messages reçus d'un processus  $P_i$  après être devenu rouge. Lorsque  $P_i$  devient rouge, il transmet un *marker* à  $P_j$  qui ne mémorise plus les messages.



# Algorithme de Chandy et Lamport

```
public class RecvCamera extends Process implements Camera {  
    static final int white = 0, red = 1;  
    int myColor = white;  
    boolean closed[];  
    CamUser app;  
    LinkedList chan[] = null;  
    public RecvCamera(Linker initComm, CamUser app) {  
        super(initComm);  
        closed = new boolean[N]; // on arrête de mémoriser les messages oui/non  
        chan = new LinkedList[N]; // mémorise l'état du canal  
        for (int i = 0; i < N; i++)  
            if (isNeighbor(i)) {  
                closed[i] = false; // mémorisation  
                chan[i] = new LinkedList();  
            }  
        else closed[i] = true;  
        this.app = app;  
    }  
}
```

# Algorithme de Chandy et Lamport

```
public synchronized void globalState() {  
    myColor = red;  
    app.localState();    // On mémorise l'état local  
    sendToNeighbors("marker", myId);    // on transmet le marker  
}
```

```
boolean isDone() {    // test si l'algorithme est terminé  
    if (myColor == white) return false;  
    for (int i = 0; i < N; i++)  
        if (!closed[i]) return false;  
    return true;  
}
```

# Algorithme de Chandy et Lamport

```
public synchronized void handleMsg(Msg m, int src, String tag) {  
    if (tag.equals("marker")) {  
        if (myColor == white) globalState(); // changement de couleur  
        closed[src] = true; // on arrête de mémoriser les messages reçus  
        if (isDone()){ // l'algorithme est terminé  
            System.out.println("Channel State: Transit Messages ");  
            for (int i = 0; i < N; i++)  
                if (isNeighbor(i)) // affiche l'état des canaux  
                    while (!chan[i].isEmpty())  
                        System.out.println( ((Msg) chan[i].removeFirst()).toString());  
        }  
    } else { // message des applications, on mémorise  
        if ((myColor == red) && (!closed[src]));  
            chan[src].add(m); app.handleMsg(m, src, tag);  
            // on transmet le message à l'application concernée  
        }  
    }  
}
```

# Canaux non FIFO

Si les canaux de transmission ne sont plus FIFO, le *marker* ne peut pas être utilisé tel quel pour indiquer la fin de la mémorisation des messages. De plus, on ne peut plus se contenter de transmettre une seule fois un *marker* car il peut arriver à un processus après des messages transmis plus tard.

La couleur d'un processus est incluse dans tous les messages, ainsi un processus peut déterminer si un message a été transmis avant ou après que le processus expéditeur ait changé de couleur.

Dans le *marker*, le processus expéditeur inclut le nombre de messages blanc préalablement transmis, permettant à un processus de déterminer

# Impossibilité

Dans les réseaux asynchrones on ne peut pas résoudre le problème du consensus par un protocole (algorithme) qui tolère la panne d'un processus (au moins). Ce résultat peut être étendu considérablement.

On considère les *tâches de décision* qui déterminent une application de l'ensemble des états initiaux du système  $I$  dans un espace de décision  $D$ . A chaque état initiale, l'application associe un sous-ensemble de  $D$  qui correspond aux décisions valides.

Un tel protocole tolère une panne d'un processus si

1. le protocole s'exécute conformément à la spécification si tous les processus sont corrects
2. si un processus ne s'exécute pas correctement, les autres processus terminent leur exécution

# Modèle

On considère un réseau dans lequel l'identité de tous les processus est connue et le graphe de communication est complet.

Les canaux de transmissions sont supposés fiables, les messages transmis arrivent avec un délai fini mais non borné (asynchrone).

**définition:** une tâche de décision  $T$  distribuée est une application

$$T : I^N \rightarrow 2^D$$

où  $N$  est le nombre de processus,  $I$  est l'ensemble des états initiaux des processus et  $D$  l'ensemble des états finaux.

# Modèle

Un vecteur d'entrée  $x$  est un N-uple  $(x_1; x_2; \dots; x_N)$  où  $x_i$  est l'état initial du processus  $p_i$ .

Un vecteur de décision  $d$  est un N-uple  $(d_1; \dots; d_N)$  où  $d_i$  est l'état final du processus  $p_i$ .

On note  $D_T$  l'ensemble des vecteurs décision.

**Exemple:** Pour le problème du consensus on a  

$$D_T = f(0; 0; \dots; 0); (1; 1; \dots; 1)g$$

Pour le problème de l'élection on a  

$$D_T = f(1; 0; \dots; 0); (0; 1; 0; \dots; 0); \dots; (0; \dots; 0; 1)g$$

# décisions adjacentes

**définition:** deux vecteur de décision  $d_1$  et  $d_2$  sont adjacents si ils sont différents pour une seule composante.

**définition:** le graphe de décision d'une tâche  $T$  à pour sommets les éléments de  $D_T$  et pour arêtes

$$E_T = \{ (d_1; d_2) : d_1; d_2 \text{ sont adjacents } g \}$$

**définition:** une tâche est non connexe si sont graphe de décision est non connexe.



# Décisions partielles

Lorsque le protocole (algorithme distribué) s'exécute les processus corrects vont déterminer une valeur de décision. A un instant donné, il se peut qu'un sous-ensemble de processus n'ait pas encore pris de décision. Le vecteur  $(d_1, \dots, d_N)$  est un vecteur de *décision partiel*. On note les composantes de ce vecteurs pas encore définitive B-composantes.

Soit  $d_1$  un vecteur de décisions,  $d_2$  un vecteur de décision partiels est cohérent avec  $d_1$  si ce dernier peut être obtenu en modifiant les B-composantes de  $d_2$ .

**définition:** un protocole P résout une tâche T si pour chaque vecteur  $d_2 \in D_T$  il existe une exécution correspondante.

# Résultat

**Lemme:** Soit  $T$  une tâche de décision et  $P$  un protocole qui résout la tâche et tolère un processus non correct.

Une exécution ou un processus est non correct s'arrête dans un état  $C$  et **les processus corrects déterminent un vecteur de décisions partiel qui est cohérent avec un vecteur de décision.**

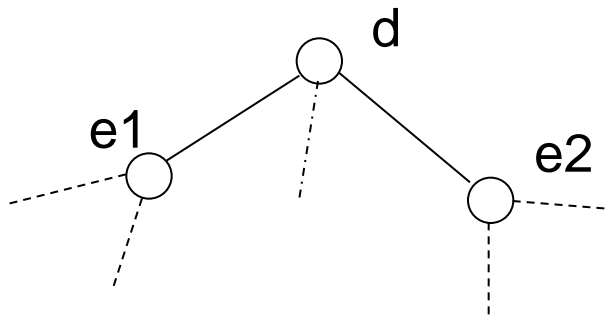
**Preuve:** Supposons que le vecteur soit incohérent avec tous les vecteurs de décisions. Il se peut que le processus non correct soit seulement très lent... Après que  $N-1$  processus aient pris leur décision et déterminé le vecteur de décisions partiel il commence à s'exécuter.

**L'exécution finale détermine un vecteur de décision et tous les processus sont corrects.**

# Résultat

**Lemme:** Soit  $d$  un vecteur de décision,  $N-1$  composantes de  $d$  déterminent la composante connexe de  $d$ .

**preuve:** On considère deux sous-ensemble de  $N-1$  composante de  $d$ , ils déterminent deux vecteur de décisions partiels  $d_1$  et  $d_2$ .  
Ces deux vecteurs sont cohérents avec les vecteurs de décisions  $e_1$  et  $e_2$  qui sont tous les deux adjacents à  $d$ .



# Finalement

**Théorème:** Une tâche de décision non connexe ne tolère pas de faute.

**preuve:** On note  $C_1, C_2, \dots, C_k$  les composantes connexes du graphe de décision. On montre que s'il existe un protocole  $P$  pour résoudre  $T$  tolérant alors il existe un protocole  $P'$  tolérant pour résoudre le problème du consensus.

Les processus exécutent le protocole  $P$ . Chaque processus qui prend une décision transmet la valeur à tous les processus. On suppose qu'un processus est non correct.

Au bout d'un temps fini, les  $N-1$  processus corrects connaissent le vecteur de décisions partiel. Ils déterminent la composante connexe du vecteur de décision cohérent (les processus sont identifiés!). Tous les processus déterminent la même composante connexe et donc on a consensus sur la valeur choisie, **une contradiction.**