

## ETAT GLOBAL D'UN SYSTEME REPARTI

- **ETAT LOCAL  $E_{Li}$  D'UN SITE  $S_i$**   
état initial et séquence d'événements locaux sur  $S_i$
- **ETAT LOCAL  $E_{Cij}$  D'UN CANAL  $C_{ij}$**   
ensemble des messages en transit sur le canal  $C_{ij}$   
émis par  $S_i$  et non encore reçus par  $S_j$
- **EVENEMENTS (ATOMIQUES) FAISANT EVOLUER LE SYSTEME**  
 $\{E_{Li}\}$  événement interne sur  $S_i$   $\{EL'i\}$   
 $\{E_{Li}, E_{Cij}\}$  émission de  $m$  par  $S_i$  sur  $C_{ij}$   $\{EL'i, EC'ij = E_{Cij} \cup \{m\}\}$   
 $\{E_{Li}, E_{Cki}\}$  réception de  $m$  par  $S_i$  sur  $C_{ki}$   $\{EL'i, EC'ki = E_{Cij} - \{m\}\}$
- **ETAT GLOBAL  $S = \{ \text{pour tout } i, j, \cup E_{Li}, \cup E_{Cij} \}$**
- **DIFFICULTES**  
 $E_{Li}$  n'est immédiatement observable que sur  $S_i$   
 $E_{Cij}$  n'est jamais directement observable, ni sur  $S_i$ , ni sur  $S_j$
- **OBJECTIF : définir un état global déterminable localement par tout site**

**REMARQUE : toute définition d'état doit respecter la dépendance causale**

**nota : on dit indifféremment site ou processus  
(dans ce cas il y a un processus par site)**

## PASSÉ D'UN ÉVÉNEMENT

**Le passé (ou historique) d'un événement  $e$ , c'est par définition :**

$$\text{hist}(e) = e \cup \text{ensemble des événements } e' \text{ tels que } e' \rightarrow e$$

**Seul le passé de  $e$  peut avoir influencé  $e$**

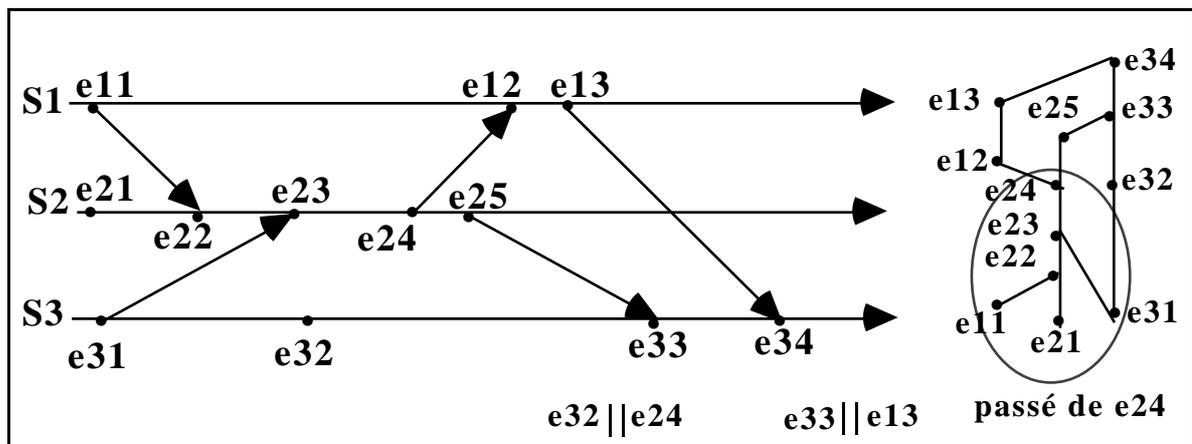
**Chaîne causale :  $e_0 \dots e_n$  tels que  $e_{i-1} \rightarrow e_i$  pour tout  $i \in (1..n)$**

**Événements concurrents (ou encore causalement indépendants)**

$$a \parallel b \Leftrightarrow \text{non } (a \rightarrow b) \text{ et non } (b \rightarrow a)$$

**aucun des 2 événements n'appartient au passé de l'autre**

**aucun des 2 événements ne peut influencer l'autre**



### Applications

**définition de la cohérence d'un état, d'une observation**

**mise au point répartie**

**mesure du parallélisme**

# COUPURES

soit  $E$  un ensemble d'événements constituant une application répartie

**Coupure** = sous-ensemble fini  $C$  de  $E$  tel que pour  $a, b \in E$

$$a \in C \text{ et } (b \text{ précède localement } a) \Rightarrow (b \in C)$$

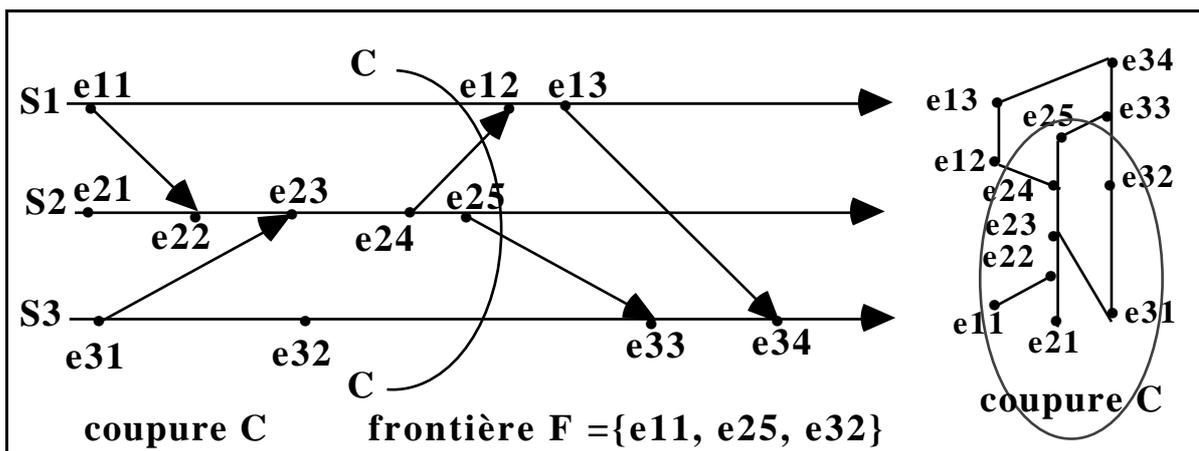
**Photographie instantanée d'un système** obtenue en prenant un événement par site et tous les événements du site qui le précède.

$C$  est un sous-ensemble de l'histoire de l'application qui contient toute l'histoire qui le précède : cela permet de définir un passé et un futur (par rapport à la coupure)

**Frontière  $F$  d'une coupure  $C$  :**

ensemble des événements les plus récents de la coupure, un par site

$$a \in F \Leftrightarrow a \in C \text{ et il n'existe pas de } b \in C \text{ tel que } a \rightarrow b$$



## COUPURES COHERENTES

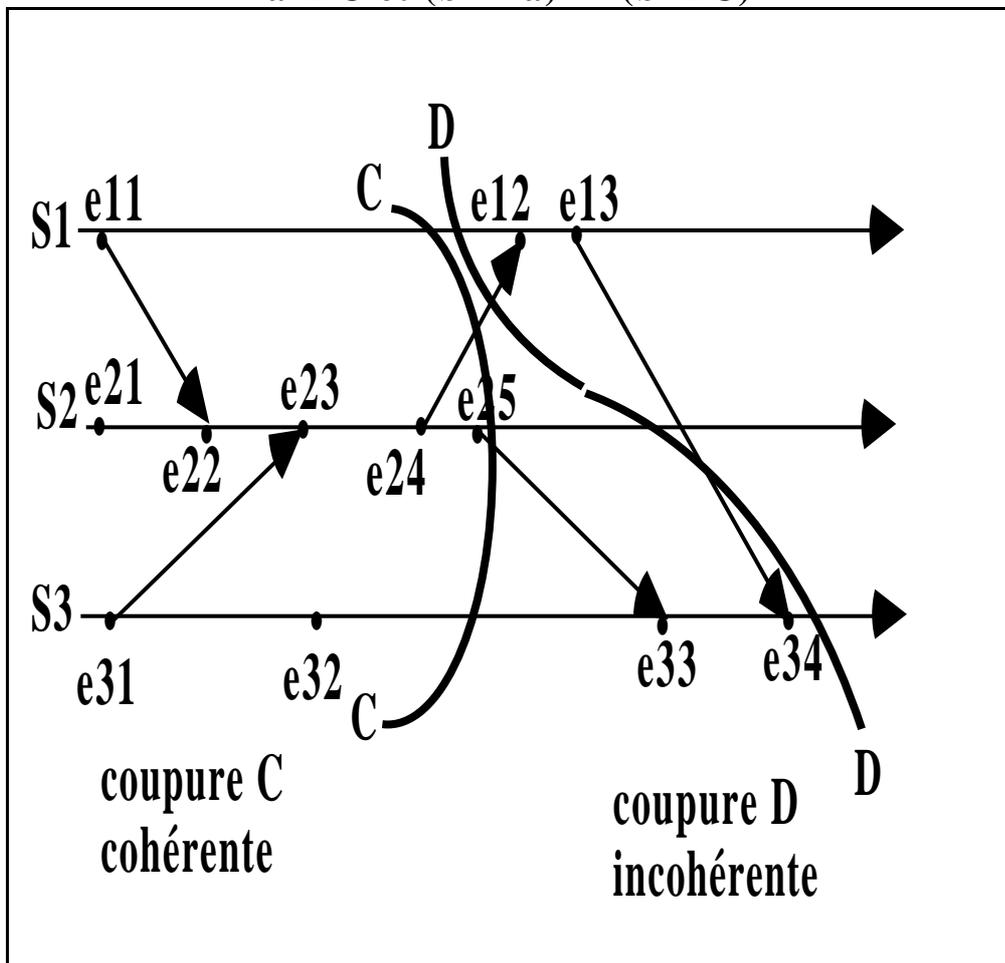
**Cohérence = respect de la causalité dans la coupure**

**une chaîne causale ne peut sortir et re-entrer dans la coupure**

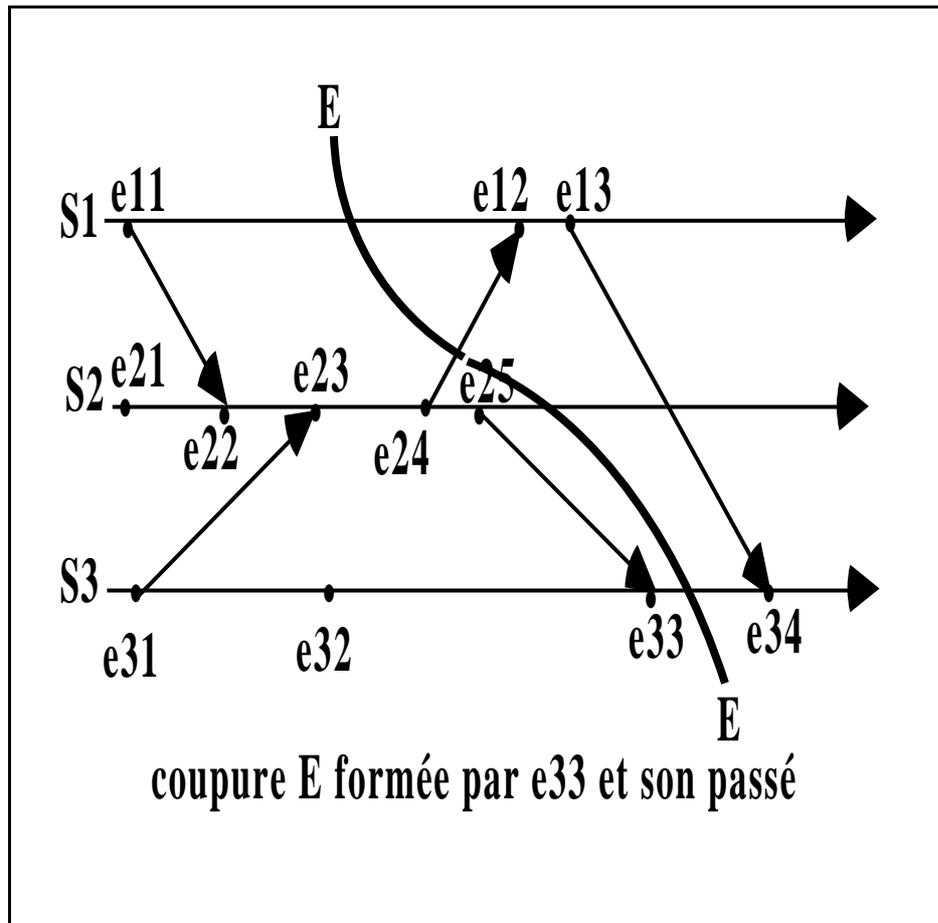
**un message ne peut pas venir du futur en franchissant la frontière**

**Coupure cohérente = coupure fermée par la relation de dépendance causale**

$$a \in C \text{ et } (b \rightarrow a) \Rightarrow (b \in C)$$



**Etat global cohérent = état associé à une coupure cohérente**



Exemple de coupure cohérente : le passé d'un événement

## ETAT GLOBAL COHÉRENT D'UN SYSTEME REPARTI

- soit  $EL_i$  état local du site  $S_i$  pour tout  $i$  (histoire locale de  $S_i$ )
- soit  $EC_{ij}$  état local du canal  $C_{ij}$  pour tout  $(i, j)$

état global  $S = \{ \text{pour tout } i, j, \cup EL_i, \cup EC_{ij} \}$

coupure associée =  $\{ \text{pour tout } i, \cup EL_i \}$

état global  $S$  cohérent si coupure associée cohérente

La frontière de la coupure associée définit un passé et un futur

**Il en résulte deux conditions nécessaires pour les messages  $m$  qui traversent une frontière de coupure**

**condition C1 :**

**si  $EMISSION_i(m) \in EL_i$  alors**

**soit  $RECEPTION_j(m) \in EL_j$ , soit  $m \in EC_{ij}$**

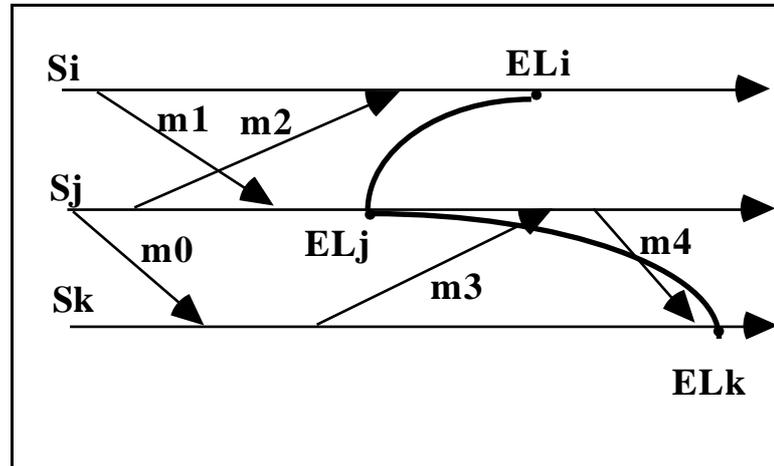
**tout message émis dans le passé est soit reçu dans le passé soit en transfert**

**condition C2 :**

**si  $EMISSION_i(m) \notin EL_i$  alors  $RECEPTION_j(m) \notin EL_j$**

**tout message émis dans le futur reste dans le futur**

## UNE EXECUTION REPARTIE



état global  $\{EL_i, EL_j, EL_k\}$

1•  $EMISSION_k(m_3) \in EL_k$  mais

comme  $RECEPTION_j(m_3) \notin EL_j$ ,

il faut que  $m_3 \in E_{Ckj}$

donc la condition C1 est vraie seulement si  $m_3 \in E_{Ckj}$

2•  $EMISSION_j(m_4) \notin EL_j$  mais  $RECEPTION_k(m_4) \in EL_k$

la condition C2 est fausse car

C2 : si  $EMISSION_j(m_4) \notin EL_j$  alors  $RECEPTION_k(m_4) \notin EL_k$

conclusion

$\{EL_i, EL_j, EL_k\}$  n'est pas un état global cohérent

# **DETERMINATION D'UN ETAT GLOBAL COHERENT**

**(Chandy - Lamport 1985)**

## **◆ Hypothèses**

- **le réseau de communication est connexe**
- **les canaux respectent l'ordre d'émission des messages (canaux FIFO)**
- **un site "élu" particulier lance la détermination d'état global**

## **◆ Principe**

- **association d'un message marqueur à chaque enregistrement d'état local d'un site pour que les autres sites puissent repérer les messages qui sont avant ou après cet enregistrement (les messages du passé et du futur)**
- **chaque site détermine son état local et celui de ses canaux en réception, puis envoie ces états au site "élu"**

## **◆ Propriétés**

- **l'algorithme se termine**
- **l'état global enregistré correspond à une coupure cohérente**
- **les états enregistrés des canaux sont corrects pour cette coupure**

## **SOLUTION DE CHANDY et LAMPORT (1985)**

### **◆ HYPOTHÈSE : CANAUX FIFO**

**tout message  $m_k$  envoyé sur un canal FIFO sépare les messages du canal en deux sous-ensembles :  $<m_k$  (avant  $m_k$ ) et  $>m_k$  (après  $m_k$ ).**

### **◆ CONTRAINTES DE COHÉRENCE**

**• Quand  $S_i$  enregistre son état  $EL_i$ , toutes les émissions de messages faites par  $S_i$  avant  $EL_i$  sont captées dans  $EL_i$  et les réceptions de ces messages doivent être captées dans les  $EL_j$  et  $C_{ij}$  des sites  $S_j$  et celles-là seulement.**

### **◆ MISE EN OEUVRE**

**• Dès que  $S_i$  enregistre  $EL_i$ , il émet un message marqueur  $m_k$  sur chaque canal  $C_{ij}$ .  $S_j$  doit enregistrer  $EL_j$  au plus tard à la réception de  $m_k$  et  $S_j$  doit capter tous les messages  $<m_k$ , soit dans  $EL_j$  soit dans  $C_{ji}$ . Les messages de  $>m_k$  ne doivent pas être captés car ils viennent du futur de  $EL_i$  sur  $S_i$ .**

### **◆ ALGORITHME SUR CHAQUE SITE $S_i$**

**◆ Début ( $S_i$  "élu") ou première réception d'un marqueur  $m_k$  (émis par  $S_j$ )**

**(1) enregistrer  $EL_i$**

**(2) enregistrer  $EC_{ji}$  comme vide car  $S_i$  a déjà reçu tous les messages  $<m_k$  et ceux de  $>m_k$  ne font pas partie de l'état global.**

**Au démarrage sur "élu", aucun  $EC_{ji}$  n'est enregistré.**

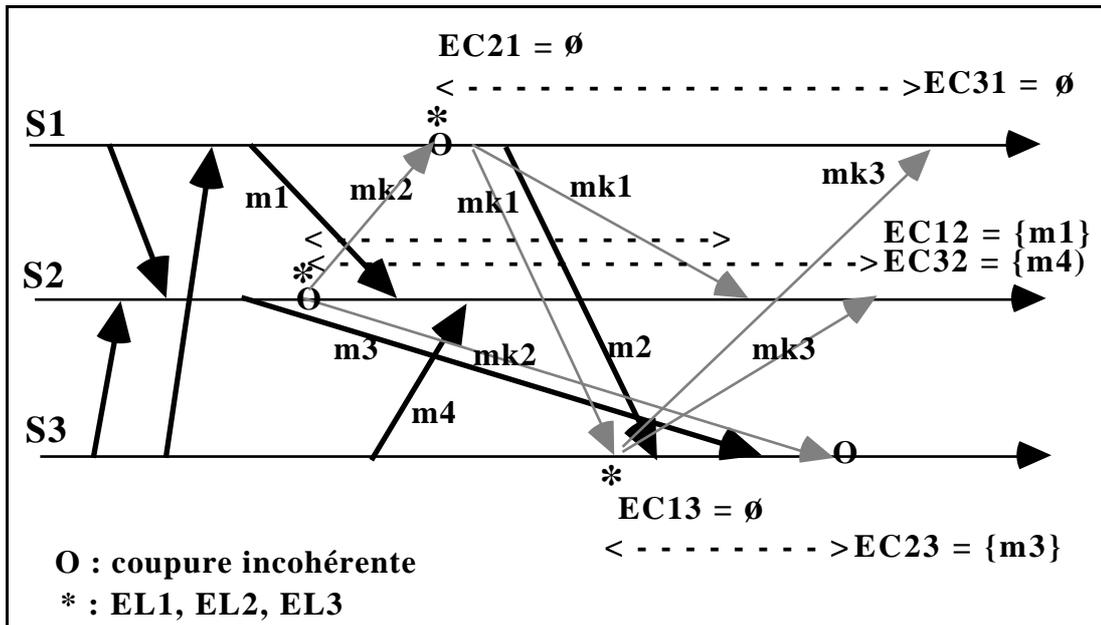
**(3) diffuser  $m_k$  à tous ses voisins sur les canaux  $C_{ij}$**

**(1-2-3) doit être atomique**

**◆ Réceptions suivantes du marqueur  $m_k$  (émis par  $S_j$ )**

**enregistrer  $EC_{ji}$  comme constitué des messages  $<m_k$  reçus par  $S_i$  entre l'enregistrement de  $EL_i$  et l'arrivée de  $m_k$  (envoyés par  $P_j$  avant  $EL_j$  mais pas reçus par  $P_i$  au moment de  $EL_i$ )**

# EXEMPLE DE DETERMINATION D'UN ETAT COHERENT

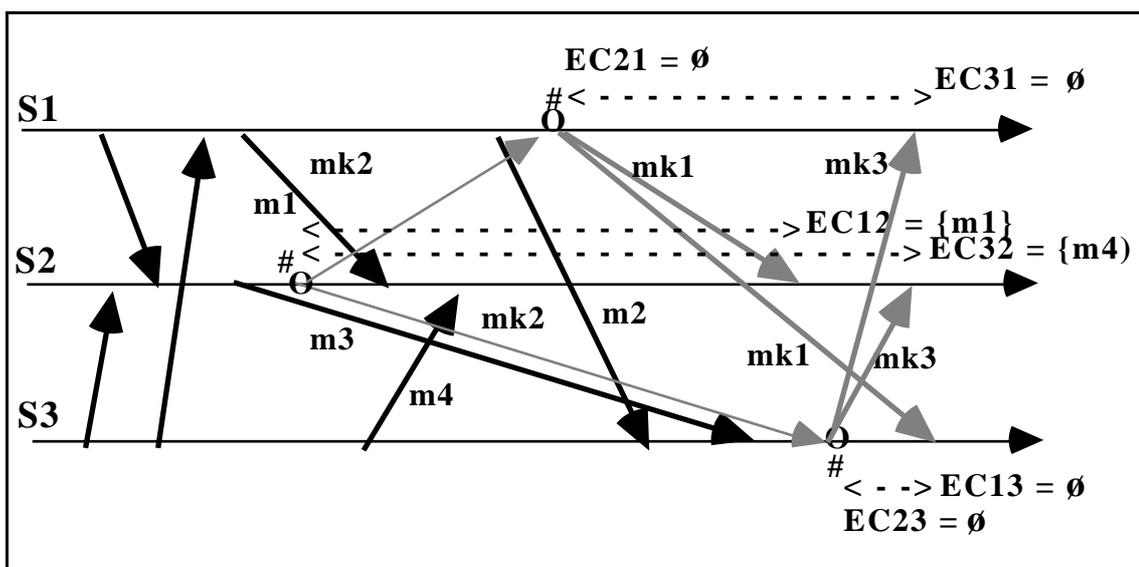


**S2 lance la détermination d'état global et diffuse mk2**

**Les 3 événements émission<sub>2</sub>(mk2), réception<sub>1</sub>(mk2) et réception<sub>3</sub>(mk2) déterminent une coupure incohérente.**

**Les marqueurs mk1 et mk3 permettent :**

- **de forcer la transitivité de la causalité et d'avoir une coupure cohérente.**
- **de noter dans les EC<sub>ij</sub> les messages qui traversent la coupure cohérente.**



**autre trace et autre état cohérent**

## **DETERMINER UN ETAT GLOBAL DANS UN SYSTEME REPARTI**

### **éléments de bibliographie**

#### **Solution pour un canal FIFO :**

**K.M. Chandy and L.Lamport, *Distributed Snapshots : Determining Global States of Distributed Systems*. ACM TOCS, Vol 3,1, (1985) pp. 63-75**

#### **solutions pour les canaux non FIFO**

**méthode cumulative (et rapide) de Lai et Yang : T.H.Lai and T.H.Yang, *On Distributed Snapshots*; Inf. Proc. Letters, Vol. 25, (1987), pp. 153-158**

**méthode non cumulative (mais lente) de Mattern : F.Mattern, *Virtual Time and Global States of Distributed Systems*. Proc. of Int. Workshop on Parallel and Dist. Systems, North Holland, 1988, pp. 215-226**

**solution utilisant des messages de contrôle : M.Ajuha, *Global Snapshots for Asynchronous Distributed Systems with non FIFO Channels*. Tec. Rep. #CS92-268, U. of Calif., San Diego (1992), 7 p.**

#### **solutions fondées sur l'ordre causal**

**méthode centralisée d'Acharya et Badrinath : A.Acharya and B.R.Badrinath, *Recording Distributed Snapshots Based on Causal Order of Message Delivery*. Inf. Proc. Letters, Vol. 44, (1992), pp.317-321**

**méthode répartie d'Alagar et Venkatesan : S.Alagar and S.Venkatesan, *An Optimal Algorithm for Distributed Snapshots with Causal Message Ordering*, Tec. Rep, U. of Texas, Dallas (1993), 7 p.**

#### **DOCUMENTATION UTILISEE**

**J.M.Helary, A.Mostefaoui, M.Raynal, *Déterminer un état global dans un système réparti*, RR. 2090, INRIA (1993), 21 p.**