

# Protocoles de diffusion atomique



Institut de Recherche en Informatique de Toulouse

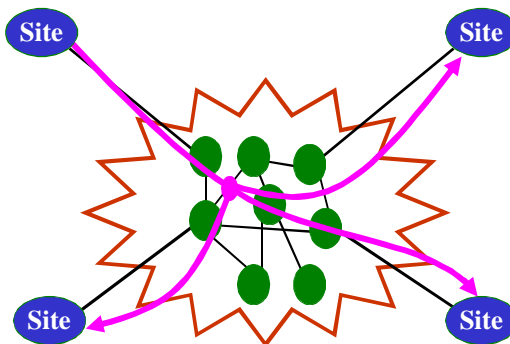
Z. Mammeri  
IRIT - UPS

Cours de DEA, 1998-2004



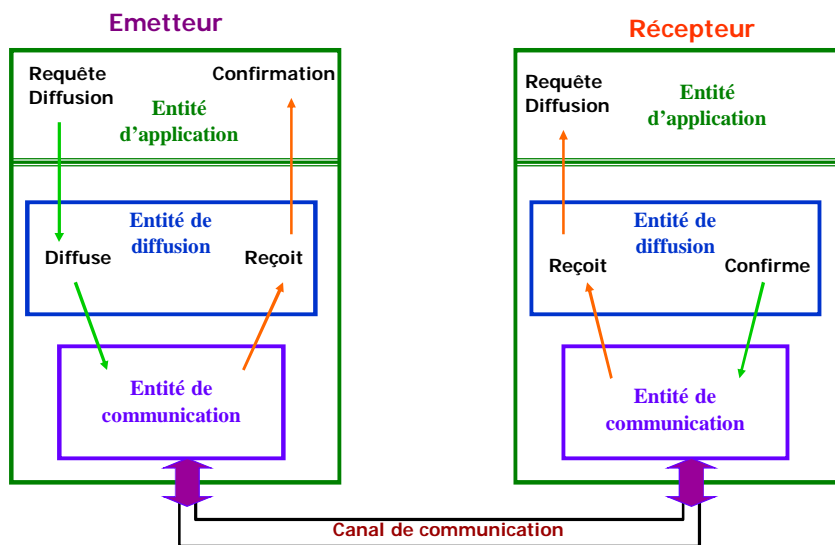
## 1. Problèmes et propriétés

- **Groupe** = ensemble de sites (ex. participants à une téléconférence)
- **Service de diffusion** au sein d'un groupe



- **Problème** : comment garantir que tous les sites corrects reçoivent les mêmes messages et dans le même ordre ?

## Entités impliquées dans la diffusion



## Propriétés attendues des protocoles de diffusion

- **Consensus unanime** (c'est le consensus le plus souvent requis)  
Si un site diffuse un message, alors soit tous les sites corrects remettent ce message à leurs entités d'application, soit aucun d'entre eux ne le remet.
- **Consensus uniforme**  
Si un site correct ou incorrect remet un message à son entité d'application, alors tous les sites corrects remettent ce message à leurs entités d'application.
- **Terminaison**  
Chaque site correct connaît l'issue (fin) de la diffusion en un temps fini.
- **Ordre total**  
Tous les sites corrects remettent les messages dans le même ordre à leurs entités d'application.

## Structures et hypothèses d'implantation des protocoles de diffusion (1)

- **Modèles de systèmes répartis**

- Synchrones (avec délai de communication borné)
- Asynchrones (avec délai de communication non borné)  
⇒ **Pas de garantie de terminaison**

- **Type de contrôle**

- Centralisé
- Décentralisé

- **Type de validation**

- Symétrique : la décision de valider un message diffusé est prise symétriquement par l'émetteur et par les récepteurs, chacun décide de valider ou non selon les informations dont il dispose.
- Asymétrique : l'émetteur seul est chargé de collecter les acquittements qui lui permettent de décider de la validation des messages. La décision qu'il prend est alors communiquée aux récepteurs.

## Structures et hypothèses d'implantation des protocoles de diffusion (2)

- **Défaillance**

- Absence de défaillance
- Défaillance de : l'émetteur? du récepteur? du réseau? ... des trois?

- **Types de fautes**

- Fautes par omission
- Fautes temporelles (messages arrivant en retard)
- Fautes byzantines

- **Fonctionnement après la faute**

- Arrêt sur défaillance (oui/non)
- Isolation après défaillance (oui/non)

- **Annulation de message après la faute**

- Annuler le message dès que l'émetteur devient défaillant
- Annuler le message si aucun site ne peut poursuivre la diffusion après la défaillance l'émetteur initial.

## Structures et hypothèses d'implantation des protocoles de diffusion (3)

- **Diffusion spontanée (autorisée ou non)**

Un émetteur sait à quel moment il a le droit d'émettre son message ou bien il peut le faire à n'importe quel moment (diffusion spontanée).

- **Diffusions concurrentes**

- Un seul message diffusé à la fois dans un même groupe.
- Plusieurs messages peuvent être diffusés simultanément dans un même groupe.

- **Exclusion des membres du groupe**

- Si l'émetteur ne reçoit pas d'acquittement d'un récepteur, il l'exclut du groupe.
- L'exclusion d'un membre résulte d'un consensus majoritaire au sein du groupe.

- **Accès au temps**

- Horloges locales non synchronisées
- Horloges locales synchronisées
- Horloge globale unique

## Structures et hypothèses d'implantation des protocoles de diffusion (4)

Beaucoup de choix sont possibles.

Beaucoup de combinaisons de fonctions sont possibles.

Beaucoup d'hypothèses sont possibles.



Conséquences sur la complexité des solutions pour la diffusion atomique



Beaucoup de protocoles ont été proposés dans la littérature.

## 2. Exemple 1

### Protocole à contrôle centralisé [Tanenbaum]

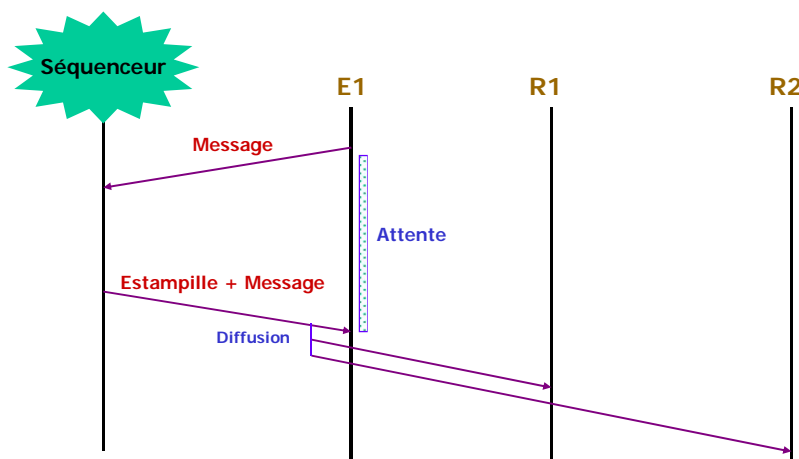
- **Hypothèses**

- Les sites sont reliés par un bus à diffusion
- Les délais de communication sont bornés
- Fautes autorisées : omissions et fautes temporelles
- Mécanisme de redondance pour éviter le partitionnement du système
- Chaque site accède à son horloge locale seulement

- **Propriétés : le protocole garantit**

- consensus unanime
- terminaison
- ordre total

- **Principe du protocole**



## ● Description du protocole

1. Pour diffuser son message, l'émetteur l'envoie au séquenceur.
2. Le séquenceur attribue un numéro d'ordre (une estampille) au message et le diffuse aux membres du groupe.
3. L'émetteur se bloque jusqu'à ce qu'il reçoive l'estampille attribuée au message. Un temporisateur est utilisé pour retransmettre le message au séquenceur (l'émetteur peut soumettre plusieurs fois son message).
4. Le séquenceur conserve une copie des messages qu'il a diffusés avec leur estampille dans un tampon d'historique. Si tous les acquittements relatifs à un message ont été reçus, ce message est retiré du tampon d'historique. Il s'agit donc d'une validation asymétrique.
5. Chaque destinataire accepte les messages dans l'ordre de leur estampille et les délivre dans cet ordre. Il conserve le numéro du dernier message reçu, ce qui permet, le cas échéant, de demander au séquenceur de lui envoyer les messages manquants.

## 7. Protocole de synchronisation à deux phases :

**Phase 1 :** le séquenceur diffuse un message d'invitation contenant le dernier numéro d'estampille utilisé. Sur réception de ce message, chaque récepteur vérifie qu'il ne lui manque pas de messages. S'il est à jour, il acquitte le séquenceur, sinon il lui demande tous les messages manquants.

A la fin de la phase 1, tous les récepteurs ont reçu tous les messages diffusés, permettant au séquenceur de vider son tampon d'historique.

**Phase 2 :** le séquenceur diffuse un message de validation informant tous les membres du groupe que tous les sites sont à jour. Sur réception d'un message de validation, chaque récepteur envoie un acquittement de validation et reprend le cours normal du protocole de diffusion. Le séquenceur reprend le cours normal du protocole de diffusion quand il a reçu tous les acquittements attendus.

8. Pendant la phase de synchronisation, le séquenceur ignore tous les nouveaux messages qui lui sont soumis par les sites.

### 9. Défaillance d'un site

Après avoir reçu  $X$  messages, chaque site envoie au séquenceur un numéro de vie indiquant le numéro d'estampille du dernier message qu'il a reçu. L'absence du message de vie permet de détecter la défaillance de site. Ce site est alors exclu du groupe.

### 10. Défaillance du séquenceur

Si un site ne reçoit pas de réponse du séquenceur après  $Y$  tentatives de retransmission, le site suppose que le séquenceur est défaillant. Ensuite, un protocole d'élection d'un nouveau séquenceur est exécuté (ce protocole doit garantir l'unicité du séquenceur élu).

## 3. Exemple 2

### Protocole à contrôle décentralisé [Verissimo]

#### ● Hypothèses

- Les sites sont reliés par un bus à diffusion
- Le bus est sujet à perte de messages
- Fautes autorisées : omissions et fautes temporelles.
- Un site est supposé arrêté après  $w$  omissions successives
- Mécanisme de redondance pour éviter le partitionnement du système
- Chaque site accède à son horloge locale seulement

#### ● Propriétés : le protocole garantit

- consensus unanime
- terminaison
- ordre total

## ● Description du protocole

### 1. Première phase (dissémination du message):

- i– L'émetteur diffuse son message estampillé au groupe
- ii– Si à la fin de la première phase, il lui manque un acquittement, il recommence une nouvelle retransmission.
- iii– La 1ère phase s'achève si l'une des deux conditions suivantes est satisfaite :
  - \* Cas 1 : l'émetteur a reçu tous les acquittements à la fin de la même retransmission
  - \* Cas 2 : l'émetteur a atteint le nombre maximum de retransmissions.

### 2. Deuxième phase (décision):

- i– Si l'émetteur reçoit au moins un acquittement « Refus », il annule son message sinon il le valide.
- ii– S'il lui manque un acquittement, l'émetteur fait appel au **Superviseur de groupe** qui est chargé de gérer la vue du groupe.
- iii– Un message de validation n'est transmis qu'une seule fois, contrairement au message d'annulation qui doit être acquitté par tous les récepteurs.
- iv– Sur réception d'un message de validation, un site délivre le message à sa couche application, s'il a déjà reçu tous les messages d'estampille inférieure.

### 3. Défaillance d'un récepteur

Lorsqu'un émetteur détecte la défaillance d'un récepteur, il signale à son Superviseur de groupe l'identificateur du site défaillant.

### 4. Défaillance d'un émetteur

Si un récepteur n'a pas reçu de message de validation, au bout d'un délai fixé, il signale à son superviseur l'identificateur de l'émetteur défaillant ainsi que l'estampille du message non validé.



#### 5. Fonctionnement du Superviseur de groupe (SG):

- i- Le SG est un service disponible sur chaque site, mais à un instant donné un seul SG est actif.
- ii- Lorsque le SG est appelé, il suspend le trafic en cours, et ne le rétablit que lorsque le recouvrement est effectué.
- iii- La procédure de recouvrement est la suivante :
  - \* Le SG interroge les sites pour qu'ils lui fournissent l'état du dernier message reçu provenant du site défaillant.
  - \* Le SG diffuse la nouvelle vue du groupe en excluant le site défaillant et les décisions qu'il a prises concernant les messages pendants : si le SG est appelé suite à la défaillance d'un émetteur, il annule le message pendant de cet émetteur, si aucun processeur n'a reçu de validation.

#### 6. Défaillance du Superviseur de groupe

Lorsque le site supportant le SG défaille, il faut éviter que la suspension du trafic n'entraîne le blocage du système. Le problème est résolu en utilisant des timers pour détecter la défaillance du SG. Dès qu'un site détecte la défaillance du SG, il lance la procédure d'élection d'un nouveau SG et une réinitialisation de la procédure de recouvrement.

## 4. Conclusion

- La diffusion atomique est un besoin de beaucoup d'applications réparties
- Beaucoup de possibilités de conception de solution au problème de DA
- Le problème de la DA a attiré (et attire encore) les chercheurs en algorithmique distribuée.