

Synchronisation d'horloges



Institut de Recherche en Informatique de Toulouse

Z. Mammeri
IRIT - UPS

Cours de DEA, 1998-2004

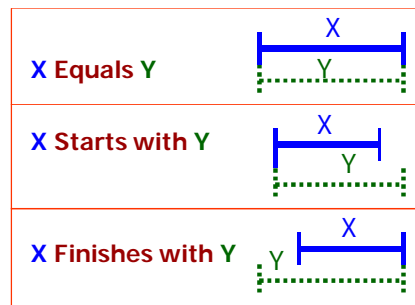
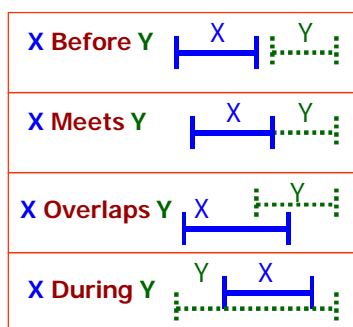


1. Temps en informatique

Temps logique vs Temps physique (*temps réel*)

Temps logique

- Opérateurs Avant et Après de Lamport (systèmes répartis)
- 7 opérateurs d'Allen



Utilisation du temps physique en informatique

- Fournir la date et heure
- Datation d'événements, Ordre des événements
- Activation de processus (tâches, threads, agents...)
 - à instant fixe
 - périodiquement
 - sur détection d'événement
 - avec un temps de réponse fixé
 - dans un intervalle de temps donné
 - ...

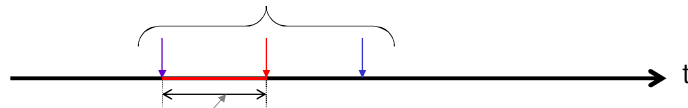
Applications ayant besoin de temps précis

- Transactions boursières
- Contrôle de trafic aérien et position d'aéronefs
- Localisation de mobiles (GPS et autres)
- Diffusion et contrôle de programmes de radio et TV
- Gestion de risques climatiques, écologiques, pollution...
- Synchronisation de flux multimédia dans la téléconférence
- Simulation interactive
- Jeux distribués, Musique distribuée
- Enseignement à distance
- Systèmes de commande en temps réel de processus physiques
- Gestion de réseaux et de systèmes de télécommunications
- Transactions sur les BD
- Documents mis à jour de manière répartie (marques des m.à.j.)
- ...

Définition du temps

- « Si on ne me le demande pas, je crois savoir ce qu'est le temps. Mais si on me le demande, je ne le sais plus ». Saint Augustin (IV^e siècle)
- **Axe du temps** : **Passé, présent, futur, continuité du temps**

Points = instants d'occurrences d'événements



Segment = **intervalle** de temps ou **période** ou **durée**

Mesure du temps (1/2)

- Pour capter la notion de temps, il faut des équipements de mesure
- **Temps réel** $\xrightarrow{\text{mesure}}$ **temps physique**

Concept abstrait

- **Équipements de mesure**
 - Rotation de la terre
 - Sablier
 - Horloge mécanique (pendule)
 - Horloge à quartz
 - Horloge atomique

Mesure du temps (2/2)

- **Borne minimale de mesure du temps**
Chronon = $5,4 \cdot 10^{-44}$ secondes (limite de la mécanique quantique)
- **Borne maximale de mesure du temps**
Vie de l'univers = 10^{17} secondes
- **Meilleure approximation du temps est donnée par des horloges atomiques à base de césium avec une précision de 10^{-14}**
(dérive d'une microseconde tous les 3 ans environ)

Histoire du calendrier (1/2)

- Besoins de calendrier chez les anciens : agriculture, rituels...
- **Proposition de divers calendriers (rotation, soleil, lune,...) :**
 - Calendrier lunaire avec 13 mois de 28 jours (année = 354 jours)
 - Calendrier lunaire avec 12 mois de 29 et 30 jours alternés (année = 364 jours)
 - Calendrier lunaire des anciens Egyptiens avec 12 mois de 30 jours (année = 360 jours)
- **14ème siècle a.c.** : les chinois ont défini l'année solaire = 365,25 jours et le mois lunaire = 29,5 jours
- **46 a.c.** : adoption du calendrier julien (élaboré par l'astronome Sosigènes à la demande de Jules César) avec 365 jours/an, plus un jour intercalé tous les 4 ans. Sur les 36 premières années, par erreur, on rajoutait 1 jour tous les 3 ans (au lieu de 4).
- **4ème siècle ap.c.** : introduction de la semaine de 7 jours (par l'empereur Constantine I).

Histoire du calendrier (2/2)

- 1545 ap.c. : le calendrier julien avait accumulé 10 jours de retard sur l'année solaire
- 1582 ap.c. : adoption du calendrier **grégorien** (suite à la réforme ordonnée par le pape Grégoire XIII) : une année = 364,2425 jours (365 jours, 5 heures, 48 minutes et 46 secondes). On rajoute un jour au mois de février, toutes les années divisibles par 4 (sauf les années divisibles par 100) et les années divisibles par 400.
- **Calendrier actuel** : le 4 octobre 1582 était le dernier jour du calendrier julien suivi le lendemain du premier jour du calendrier grégorien daté 15 octobre 1582 (consternation!).
- **Autres calendriers (perse, inca, maya, hindou, juif, musulman, chinois...)**

Temps de référence et sa dissémination (1/7)

- Depuis 1955 : Le **Temps Atomique International (TAI)** est une base de temps très précise et stable fournie à l'aide de plusieurs horloges atomiques à césium implantées à travers le monde et gérées par différentes organisations nationales (en France, le bureau des poids et mesures) .
- **Exactitude du TAI** : Dérive d'environ 1 seconde par million d'années.
- **UTC (Universal Time Coordinated)** fournit la base du temps légal (et civil) utilisé partout dans le monde. L'UTC suit TAI mais diffère d'un nombre entier de secondes (en 2004, il diffère de 32 secondes) du temps atomique TAI.
- **1 janvier 1972** : UTC a remplacé GMT (Greenwich Mean Time) pour rendre la référence du temps indépendante d'un lieu. Même si aujourd'hui GMT est encore utilisé pour parler de UTC.
- **Unité de temps = la seconde**
 - 1 sec = 1/31556925,9747 de l'année tropicale (année tropicale = 365,2421987 jours)
 - 1 sec : temps nécessaire à un atome césium-133 pour accomplir 9 192 631 770 oscillations complètes.

Temps de référence et sa dissémination (2/7)

- **Correction de l'UTC (seconde intercalaire) :**

Des secondes sont rajoutées (ou supprimées, si la terre accélère, mais cela ne s'est jamais produit pour le moment) sous le contrôle de *International Earth Rotation and Reference Systems Service* pour se caler sur le rythme de rotation de la terre (en moyenne, le soleil doit se trouver au dessus du méridien de Greenwich à 12:00:00 UTC +/-0.9 secondes).

Depuis 1972, la seconde intercalaire est insérée après la seconde 23:59:59 le dernier jour de juin et/ou décembre.

Dates des dernières secondes intercalaires entre 1972 et 1999 : 30 juin 1972, 31 décembre 1972, 31 décembre 1973, 31 décembre 1974, 31 décembre 1975, 31 décembre 1976, 31 décembre 1977, 31 décembre 1978, 31 décembre 1979, 30 juin 1981, 30 juin 1982, 30 juin 1983, 30 juin 1985, 31 décembre 1987, 31 décembre 1989, 31 décembre 1990, 30 juin 1992, 30 juin 1993, 30 juin 1994, 31 décembre 1995, 30 juin 1997, 31 décembre 1998

Temps de référence et sa dissémination (3/7)

- Chaque pays dispose de moyens d'émission (radio, satellite, câble...) de temps UTC et de fréquence standards gérés par une administration nationale.

- Les périodes de diffusion du temps UTC ou des signaux pour synchroniser les fréquences sont variables.

- Les prix des équipements de réception du temps/fréquence sont très variables

- Traconex (Equipement de réception, par onde radio, avec une précision de 10 ms) : environ 1000 €

- Spectracom 8170 (Equipement de réception, par onde radio, avec une précision de 50 µs) : environ 2000 €

- Kinematics (Equipement de réception, par satellite GOES, avec une précision de 0.5 ms) : environ 6000 €

- Kinematics (Equipement de réception, par satellite GPS, avec une précision de 200 µs) : environ 12000 €

- ...

Temps de référence et sa dissémination (4/7)

Exemples d'équipements utilisés (à Delaware University - USA)



Récepteur Spectracom 8170 WWVB

Récepteur Spectracom 8183 GPS

Récepteur Spectracom 8170 WWVB

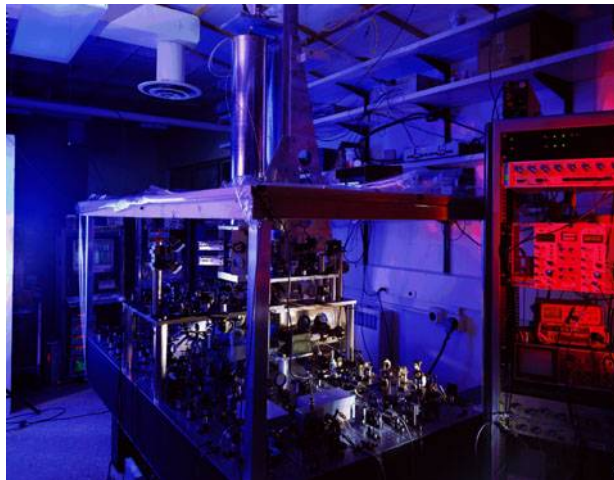
Récepteur Spectracom 8183 GPS

Quartz standard Hewlett Packard 105A

Hewlett Packard 5061A (horloge à césium)

Temps de référence et sa dissémination (5/7)

NIST-F1 Horloge (à Cesium) primaire pour l'UTC aux USA



Précision de $2 \cdot 10^{-15}$ (dérive d'une seconde pour 20 millions d'années)

Temps de référence et sa dissémination (6/7)

Temps atomique français

L'échelle de temps atomique français TA(F) est calculée mensuellement, au BNM-SYRTE (Observatoire de Paris : Bureau National de Métrologie – Systèmes de référence Temps-Espace) en utilisant les lectures quotidiennes (via le GPS) d'un ensemble d'horloges commerciales (25 en 2003) fonctionnant dans 9 laboratoires.

Les horloges des 9 laboratoires sont comparées, plusieurs fois par jour, avec l'horloge maîtresse du BNM-SYRTE.

Centre électronique de l'Armement (Rennes)	2 horloges
CNES (Toulouse)	3 horloges
France Télécom (Lannion)	2 horloges
Agilent Technologies (Palaiseau)	2 horloges
Observatoire de la Côte d'Azur (Grasse)	2 horloges
Observatoire de Besançon	3 horloges
Observatoire de Paris (BNM-SYRTE)	6 horloges
Tekelec (Les Ulis)	1 horloge
DCN Brest	4 horloges

Précision $\approx 1 \cdot 10^{-14}$

Temps de référence et sa dissémination (7/7)

Salle d'exploitation de la section Temps (Observatoire de Paris).



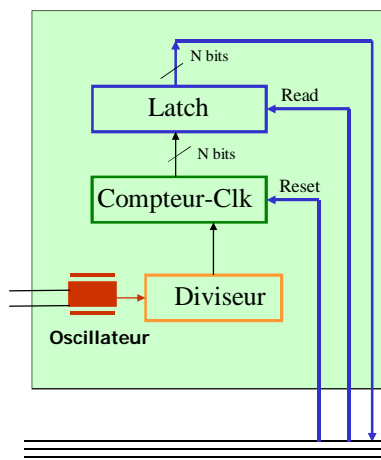
Représentation digitale de date

- En 1583 : le français Joseph Justus Scaliger a introduit la numérotation des jours pour traiter les dates comme des nombres et non sous la forme JJ/MM/AA.
- Le premier jour de l'ère julienne, noté NJJ 0.0 (Numéro de Jour Julien) a été fixé à **midi (12 h) du 1 janvier 4713 a.c.**
NJJ 1721426 correspond au 1 janvier 0001 du calendrier grégorien.
- En 1950 : introduction de la Date Julienne modifiée (DJM) pour représenter une date avec 5 chiffres. Elle a été recommandée par le CCITT et autres organismes pour son usage dans les sciences et techniques.
On traite actuellement les dates sur ordinateurs avec le MJD.
- **DJM = NJJ - 2400 000,5**
DJM = 0 correspond à 0 h du 17 novembre 1858

2. L'Horloge et ses propriétés

Horloge d'ordinateur (1/3)

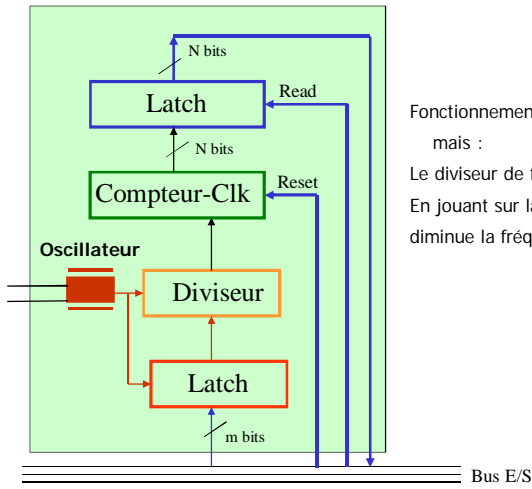
Horloge avec fréquence non modifiable



- L'oscillateur génère des impulsions avec une fréquence F
- Le diviseur génère des impulsions avec un rythme F/k
 $1/(F/k)$ = précision de l'horloge locale
- Le registre Compteur-Clk est incrémenté de 1 à chaque top envoyé par le diviseur
- Le latch permet de stabiliser le contenu du Compteur-Clk pour le lire
- La ligne Read permet de demander la lecture du registre Compteur-Clk qui est envoyé à l'utilisateur via le bus d'entrée/sortie

Horloge d'ordinateur (2/3)

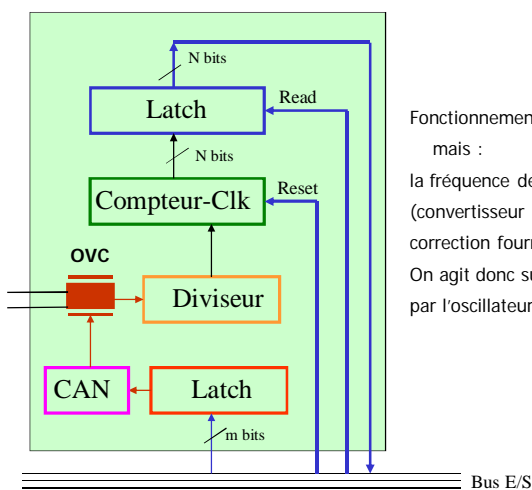
Horloge à fréquence modifiable (oscillateur non contrôlable)



Fonctionnement semblable au premier modèle d'horloge, mais :
Le diviseur de fréquence est programmable.
En jouant sur la valeur du diviseur, on augmente ou on diminue la fréquence d'incréméntation du Compteur-Clk.

Horloge d'ordinateur (3/3)

Horloge avec oscillateur contrôlable



Fonctionnement semblable au premier modèle d'horloge, mais :
la fréquence de l'oscillateur est contrôlée par un CAN (convertisseur numérique-analogique) à partir d'une correction fournie par logiciel (sous forme de m bits).
On agit donc sur le rythme de génération des impulsions par l'oscillateur.

Intervalle minimal de système

- **Intervalle minimal d'un système IM**
Si $|t(E1) - t(E2)| < IM$: alors E1 et E2 sont apparus **simultanément**
- **IM conditionne la granularité du temps requis par chaque système**
- **Valeur de IM dépend des besoins de datation de chaque système**
 - Observation des planètes
 - ...
 - Observation des mouvements des électrons

Propriétés du temps physique

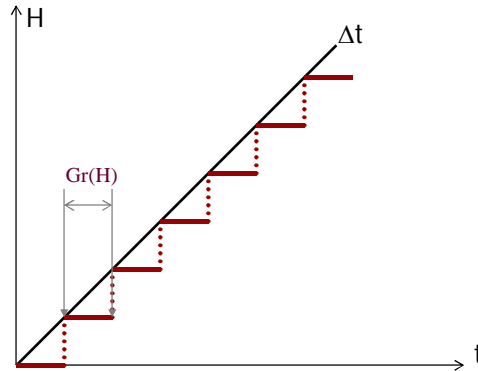
- **PTP 1** : le temps physique doit permettre de distinguer sans ambiguïté deux événements qui ne sont pas apparus simultanément si le temps les séparant est supérieur ou égal à IM.
- **PTP 2** : si un événement A est détecté avant un événement B, alors l'information temporelle (estampille) affectée à A doit être plus petite que celle affectée à B. (**La fonction d'estampillage est monotone croissante**).
- **PTP 3** : si un événement A est détecté avant un événement B, alors la différence entre les estampilles de A et B doit être approximativement égale au temps réel écoulé entre les occurrences des deux événements.
- Si le moyen de mesure du temps dont dispose un système satisfait les propriétés PTP 1, 2 et 3, alors il peut dater les événements, activer les processus, ..., sans incohérence temporelle.

Propriétés de l'horloge physique (1/3)

- **Granulat de l'horloge** = temps séparant deux tops d'horloges

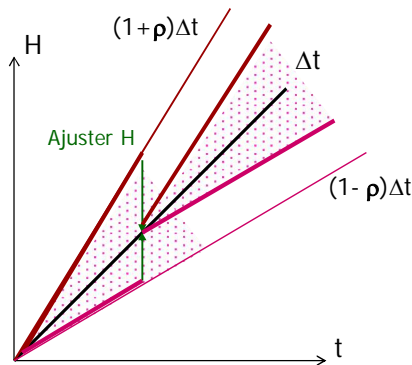
Si $|t_1 - t_2| < Gr(H)$: alors on peut avoir $H(t_1) = H(t_2)$

⇒ $H(t)$ est une fonction **continue par morceaux**



Propriétés de l'horloge physique (2/3)

- **Taux de déviation : ρ** (déviation due aux matériaux et environnement)



Condition d'horloge physique (CHP)

$$(1 - \rho) \leq \frac{H(t_2) - H(t_1)}{t_2 - t_1} \leq (1 + \rho)$$

$\rho \ll 1$: **taux de déviation de H**

$$t_2 \geq t_1 + Gr(H)$$

- **Une horloge physique qui respecte CHP est une horloge correcte**

Propriétés de l'horloge physique (3/3)

Type d'oscillateur	Dérive (en secondes)	Type de dérive
Horloge atomique à base de Maser à hydrogène	$1 \cdot 10^{-12}$ par an	Dérive constante
Horloge atomique à base de césium	$3 \cdot 10^{-13}$ par an	Dérive constante
Horloge atomique à base de rubidium	$3 \cdot 10^{-11}$ par mois	Dérive constante
Horloge à cristal à température contrôlable	$1 \cdot 10^{-10}$ par jour	Dérive variable en fonction de la température
Horloge à cristal utilisé dans les PC	$1 \cdot 10^{-9}$ par jour	Dérive variable en fonction de la température
Horloge mécanique	Seconde(s) par jour	Dérive variable en fonction de la température

Maser : Microwave Amplification by Stimulated of Radiation

Répartition et temps

- **Avantages/apports des systèmes répartis**
 - Plus de possibilités de stockage des informations
 - Plus de capacités de traitement
 - Plus de robustesse (tolérance aux fautes)
 - Plus proche de la répartition physique de l'environnement
 - ...
- **Inconvénients des systèmes répartis**
 - Sécurité/vulnérabilité
 - Pas de mémoire commune : cohérence d'état difficile
 - **Pas d'horloge commune** : différentes vues du temps
 - ...
- **Obtention d'un temps global**
 - Equipement de réception de l'UTC
 - Synchronisation d'horloges

3. Principes de base de la synchronisation d'horloges (Objectifs)

Formes de synchronisation

- Synchronisation de fréquences
Ajuster les oscillateurs pour fonctionner à une fréquence donnée
- Synchronisation de temps
Mettre à jour les horloges à une même valeur par rapport à UTC
- Synchronisation d'horloges
Synchronisation de fréquences + synchronisation de temps

Objectifs de la synchronisation d'horloges

- Fournir un temps global **approximativement** le même pour un ensemble de sites formant un "groupe", malgré:
 - les dérives des horloges physiques des sites
 - les défaillances des sites et de leurs horloges
 - les défaillances du réseau de communication
- Le temps physique fourni par une horloge synchronisée (HS) doit être une bonne approximation du temps réel pour respecter les propriétés PTP 1, 2 et 3.

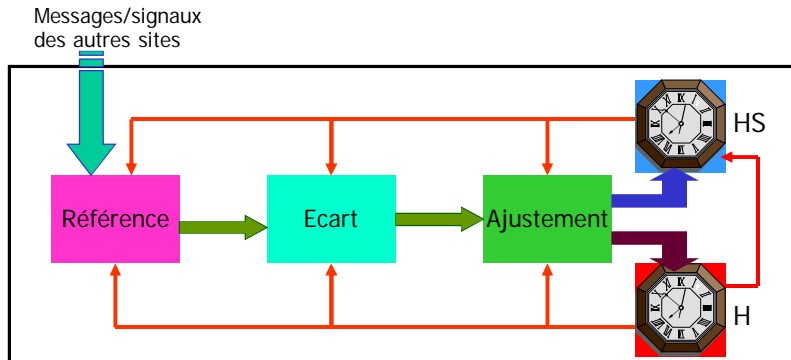
- **Propriété de précision d'accord (PA) des horloges synchronisées**

$$|HS_i(t) - HS_j(t)| \leq PA \quad 1 \leq i \leq n, \quad 1 \leq j \leq n$$

- **Propriété d'exactitude (Ex) des horloges synchronisées**

$$\exists Ex \ll 1, \quad (1 - Ex) \leq \frac{HS_i(t2) - HS_i(t1)}{t2 - t1} \leq (1 + Ex) \quad t2 \geq t1 + IM$$

Fonctions de base de la synchronisation d'horloges (1/2)

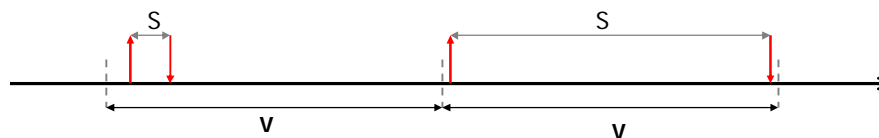


- **Référence** : calculer une valeur de référence pour ajuster l'horloge.
- **Ecart** : calculer la valeur d'écart par rapport à la référence.
- **Ajustement** : ajuster l'horloge locale.

Fonctions de base de la synchronisation d'horloges (2/2)

- **Généralement**, l'algorithme de synchronisation s'exécute avec une période V et dure S unités de temps.

V et S ont des impacts directs sur les performances de l'algorithme.



- **Arguments des fonctions**
 - valeurs d'horloges synchronisées des sites (HS)
 - valeurs d'horloges physiques des sites (H)
 - différences de valeurs d'horloges synchronisées
 - différences de valeurs d'horloges physiques

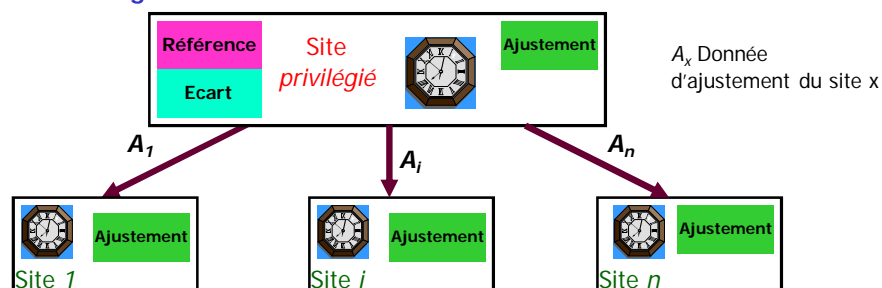
Approches de conception et de réalisation de la synchronisation d'horloges

- **Distribution des fonctions de synchronisation d'horloges**
 - Référence : Oui
 - Ecart : Oui
 - Ajustement : **Non**

- **Approches de conception**
 - Centralisée Totalement
 - Centralisée Partiellement
 - Distribuée
 - Mixte

Architecture totalement centralisée

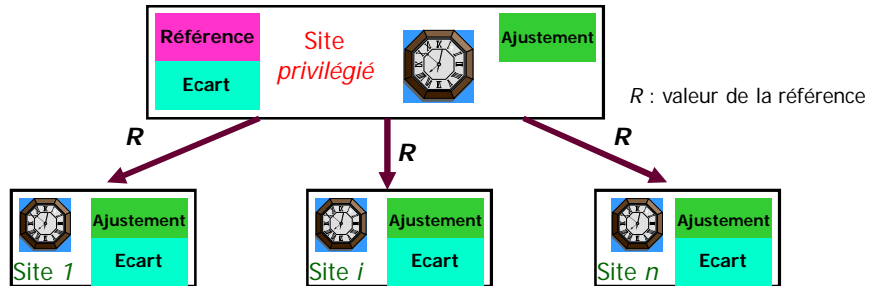
- **Un site privilégié (choisi de manière statique ou dynamique)**
 - Calcule la référence
 - Calcule les données d'ajustement de tous les sites (esclaves)
 - Envoie à chaque site esclave sa donnée d'ajustement
- **Chaque site esclave utilise la donnée d'ajustement reçue pour ajuster son horloge.**



- **Problème** : solution non tolérante aux fautes du site privilégié

Architecture partiellement centralisée

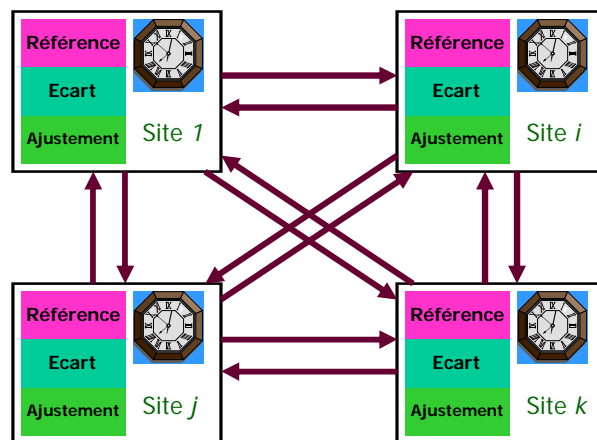
- **Un site privilégié (choisi de manière statique ou dynamique)**
 - Calcule la référence
 - Envoie à chaque site esclave la valeur de référence
- **Chaque site esclave utilise la valeur de référence reçue pour**
 - Calculer sa donnée d'ajustement
 - Ajuster son horloge



- **Problème** : solution non tolérante aux fautes du site privilégié s

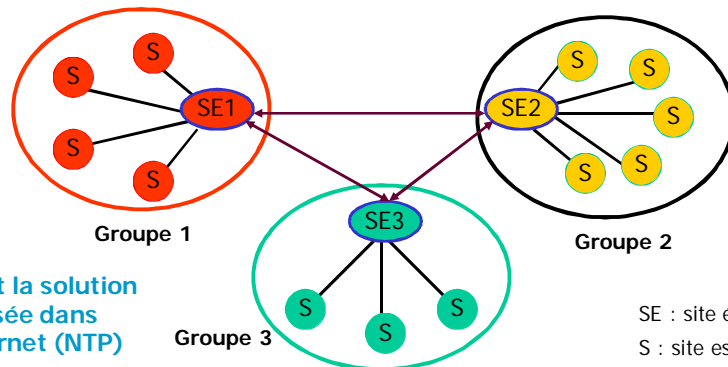
Architecture décentralisée

- **Tous les sites s'échangent des valeurs liées aux horloges et assurent les mêmes fonctions**
- **Solution tolérante aux fautes**



Architecture mixte

- **Regrouper les sites**
 - Chaque groupe est représenté par un site élu
 - Dans chaque groupe, on a une architecture centralisée
 - Les élus des groupes forment une architecture décentralisée
- **Solution utilisée pour le passage à l'échelle (scalability)**

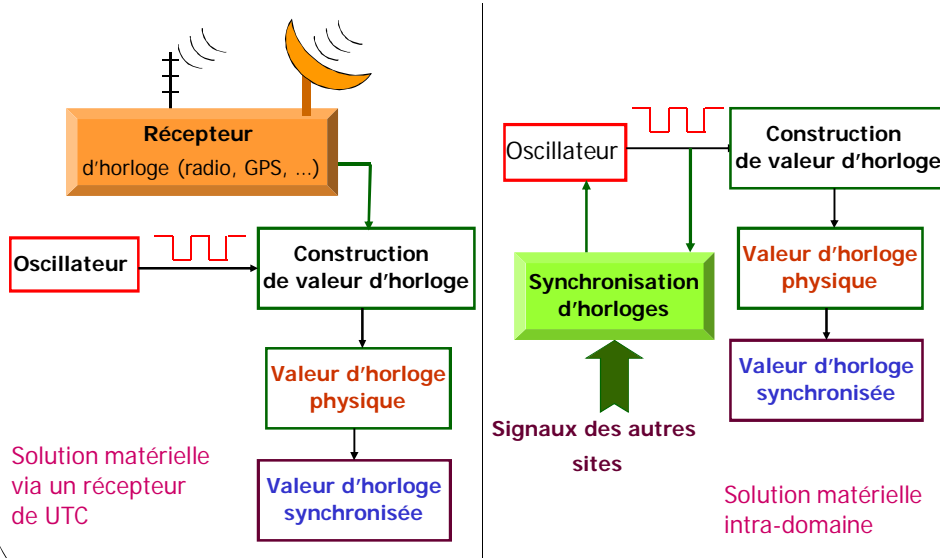


- **C'est la solution utilisée dans Internet (NTP)**

Approches de réalisation de la synchronisation d'horloges (1/3)

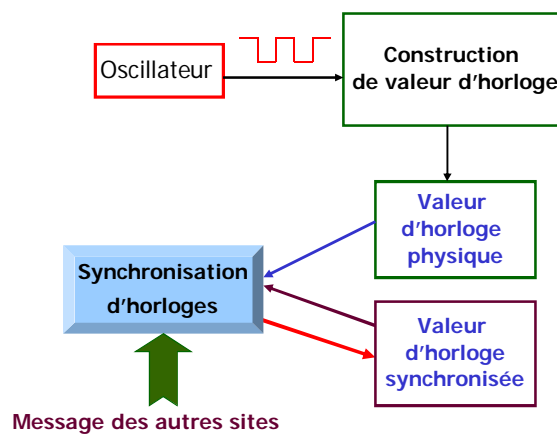
- **Approches matérielles**
 - par récepteur de temps UTC
 - par câblage et synchronisation intra-domaine
- **Approches logicielles**
- **Approches mixtes**

Approches de réalisation de la synchronisation d'horloges (2/3)



Approches de réalisation de la synchronisation d'horloges (3/3)

- **Approches logicielles**



Critères de performance de la synchronisation d'horloges

- Précision d'accord
- Exactitude de synchronisation
- Tolérance aux fautes (degré de tolérance, types de fautes...)
- Nombre de messages
- Complexité de l'algorithme (calcul)

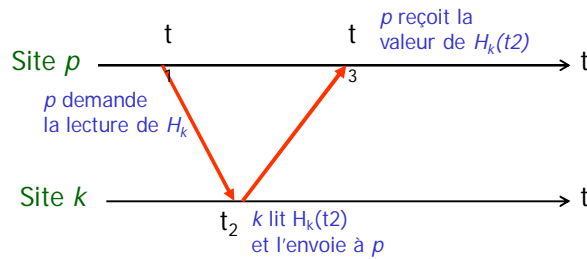
Exemple 1 : Algorithme CNV de Lamport (« Interactive CONvergence algorithm »)

- Algorithme décentralisé
- Principe de l'algorithme (pour tout site i)

```
Toutes les  $V$  unités de temps faire  
  MonHorloge =  $HS_i$   
  /* Lecture des horloges distantes */  
    Pour  $k=1$  à  $n$  et  $k \neq i$  répéter  
       $HS_k = Lire\_Horloge(k)$   
      si  $| HS_k - MonHorloge | > Seuil$   
        Alors  $HS_k = MonHorloge$   
    Fin  
  /* Calcul de la référence */  
   $R = moyenne(HS_1, HS_2, \dots, HS_n)$   
  /* Ajustement */  
   $HS_i = R$   
Fin
```

Exemple 2 : Algorithme de Gusella et Zatti

- Algorithme implanté dans le système UNIX 4.3.BSD de Berkeley
- Algorithme centralisé
- Principe de la lecture d'horloge par le site privilégié (noté site p) :



p estime la valeur de $H_k(t_3)$ par: $H_k(t_3) = H_k(t_2) + \frac{H_p(t_3) - H_p(t_1)}{2}$

- Principe de l'algorithme pour le site privilégié (noté site p)

```

Toutes les V unités de temps faire
  /* Lecture des horloges distantes */
  Pour k=1 à n et k≠p répéter
    envoi de la demande de lecture au site k ; t1,k = Hp(t)
    réception de la valeur de Hk ; t3,k = Hp(t)
  Fin
  Sélectionner les valeurs d'horloges qui diffèrent les
  unes des autres d'au maximum Z /* Z est un seuil fixé */
  somme = 0 ; NbHcorrectes = 0
  Pour k dans l'ensemble des horloges sélectionnées répéter
    Δp,k = ((t3,k + t1,k) / 2) - Hk ; /* Estimation de la différence */
    somme = somme + Δp,k ; NbHcorrectes = NbHcorrectes + 1
  Fin
  R = somme / NbHcorrectes /* Calcul de la référence */
  Pour k=1 à n et k≠p répéter
    Ak = R - Δp,k /* Calcul de la donnée d'ajustement du site k */
    envoi de Ak au site k
  Fin
  Ap = R /* Donnée d'ajustement du site privilégié */
Fin
    
```

3. Fonction Référence

Modes de calcul de la référence

- **Objectif :**
 - Calculer une **valeur de référence**
 - Déterminer un **moment d'ajustement**
- } Calcul **local** : pas d'échange de messages
 } Calcul **global** : échange de messages
- **Quatre combinaisons**
 - Calcul **global** de la valeur de référence et calcul **local** du moment d'ajustement
⇒ Algorithmes à **convergence de VALEUR**
 - Calcul **local** de la valeur de référence et calcul **global** du moment d'ajustement
⇒ Algorithmes à **convergence de MOMENT**
 - Calcul **global** de la valeur de référence et calcul **global** du moment d'ajustement
⇒ Algorithmes à **convergence de VALEUR et MOMENT**
 - Calcul **local** de la valeur de référence et calcul **local** du moment d'ajustement
⇒ **Pas de synchronisation possible**

Principe de calcul global de la référence

- **Lecture des horloges, ensuite calcul de la référence**

Valeur idéale : $R_i(tr_i) = F(X_1(tr_i), \dots, X_n(tr_i))$

Moment de calcul de la référence

Fonction de convergence

Paramètres idéaux

Paramètres estimés

$X_{ik}(t)$
= estimation de $X_k(t)$

Valeur réelle : $R_i(tr_i) = F(X_{i1}(tr_i), \dots, X_{in}(tr_i))$

- **Propriété d'invariance de la fonction de convergence**

$$F(Z_1 + Q, \dots, Z_n + Q) = F(Z_1, \dots, Z_n) + Q$$

- **Fonctions de convergence pratiques : Moyenne et Médiane**

- **Tolérance aux fautes (horloges erronées, perte de messages...)**

Calcul local de la référence ou du moment d'ajustement

- **Calcul local** ⇒ calcul fondé essentiellement sur l'horloge locale

- **Calcul local de la référence**

$$R_i(tr_i) = V * J + \beta \quad \left. \begin{array}{l} J = 1, 2, 3, \dots : n^\circ \text{ du tour de synchro} \\ V : \text{période de synchro} \\ \beta : \text{marge d'erreur} \end{array} \right\}$$

- **Calcul local du moment d'ajustement**

$$T^j = T^0 + v * J \quad \left. \begin{array}{l} J = 1, 2, 3, \dots : n^\circ \text{ du tour de synchro} \\ v : \text{période de synchro} \\ T^0 : \text{instant de début de la synchro} \end{array} \right\}$$

Condition : $\max\{T_{fin}^j - T_{début}^j\} < borne$

$T_{début}^j$: instant où la première horloge atteint $T^0 + V*j$ pour la vague j

T_{fin}^j : instant où la dernière horloge atteint $T^0 + V*j$ pour la vague j

Techniques de lecture d'horloge distante

- **Lecture d'horloges distantes = Problème n° 1 de synchro d'horloges**

- **Impossibilité d'accès immédiat à l'horloge de site distant ⇒**

Estimation des valeurs

$$ErrLect = |H_k(t) - H_{ik}(t)|$$

Erreur de lecture

Valeur réelle de l'horloge du site k

Valeur de l'horloge du k estimée par le site i

- **Objectif de technique de lecture d'horloge distante :**

Réduire l'erreur d'estimation

- **Classes de techniques de lecture d'horloge distante**

- Techniques déterministes
- Techniques probabilistes
- Techniques statistiques

Techniques de lecture déterministe (1/3)

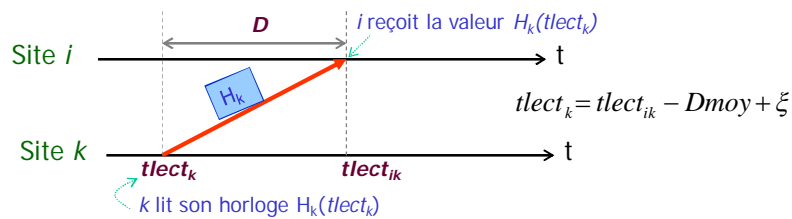
- **Déterminisme = délai de transfert de message borné**

$$D \in [D_{\min} .. D_{\max}] \quad Gig = (D_{\max} - D_{\min})$$

$$D_{moy} = (D_{\max} + D_{\min}) / 2$$

$$D = D_{moy} + \xi \quad \text{avec l'incertitude} \quad -\frac{Gig}{2} \leq \xi \leq \frac{Gig}{2}$$

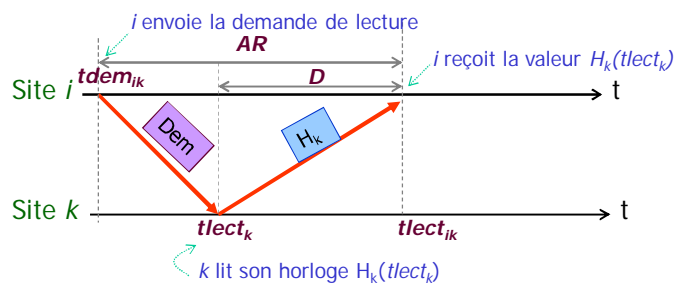
- **Technique sans demande de lecture**



Modèle d'estimation : $H_{ik}(tlect_{ik}) = H_k(tlect_k) + D_{moy}$

Techniques de lecture déterministe (2/3)

- **Technique avec demande de lecture**



- **Hypothèse I : un seul délai D_{\min} pour tous les sites**

$$tlect_k \in [t_{dem_{ik}} + D_{\min} .. tlect_{ik} - D_{\min}]$$

$$D \in [D_{\min} .. tlect_{ik} - t_{dem_{ik}} - D_{\min}]$$

$$D_{moy} = (tlect_{ik} - t_{dem_{ik}}) / 2 = AR / 2$$

Modèle d'estimation : $H_{ik}(tlect_{ik}) = H_k(tlect_k) + AR / 2$

Techniques de lecture déterministe (3/3)

- **Hypothèse II : un délai $D_{min_{ik}}$ par couple de sites et par sens**

$$tlect_k \in [t_{dem_{ik}} + D_{min_{ik}} \quad .. \quad t_{lect_{ik}} - D_{min_{ki}}]$$

$$D \in [D_{min_{ki}} \quad .. \quad t_{lect_{ik}} - t_{dem_{ik}} - D_{min_{ik}}]$$

$$D \in [D_{min_{ki}} \quad .. \quad AR - D_{min_{ik}}]$$

D peut être estimé par : $(AR + D_{min_{ki}} - D_{min_{ik}}) / 2$

Modèle d'estimation : $H_{ik}(tlect_{ik}) = H_k(tlect_k) + (AR + D_{min_{ki}} - D_{min_{ik}}) / 2$

- **Autres hypothèses → autres modèles d'estimation**

Technique de lecture probabiliste (1/2)

- Prendre en compte le **non-déterminisme** du délai de communication
- Développée à l'origine par Cristian (1989)
- **Idée de base :**

Technique de lecture avec demande + Fixer artificiellement le délai d'aller-retour en fonction de l'erreur de lecture souhaitée

Par définition de l'erreur de lecture, on a :

$$\begin{aligned} ErrLect &= |H_{ik}(tlect_{ik}) - H_k(tlect_{ik})| \\ &= |H_k(tlect_k) + AR/2 - H_k(tlect_{ik})| \end{aligned}$$

$$\text{D'où : } AR/2 \leq ErrLect - |H_k(tlect_{ik}) - H_k(tlect_k)|$$

Par définition d'une horloge correcte, on a :

$$(1 - \rho)(tlect_{ik} - tlect_k) \leq |H_k(tlect_{ik}) - H_k(tlect_k)| \leq (1 + \rho)(tlect_{ik} - tlect_k)$$

$$\text{D'où : } AR/2 \leq ErrLect - (1 - \rho)(tlect_{ik} - tlect_k)$$

$$\text{D'où : } AR/2 \leq ErrLect - (1 - \rho)D_{min}$$

Technique de lecture probabiliste (2/2)

- **Principe de la technique**

- Fixer le délai d'aller-retour, AR, pour le site k, selon l'erreur de lecture souhaitée
- Déterminer $\Pr(\text{non réception de la valeur de } H_k) = P$
 $\Pr(\text{réception de la valeur de } H_k) = 1 - P$
 Exemple. Envoyer plusieurs fois des messages et mesurer le temps d'aller-retour.
 Déduire la probabilité de dépassement de la valeur de AR choisie.
- Envoyer la demande de lecture au site k
- Si H_k n'est pas reçue au bout de AR, attendre Z unités de temps et refaire la demande.
- Si on peut tenter la lecture m fois, alors $\Pr(\text{réussite de lecture}) = 1 - P^m$
- Un choix judicieux de Z et m rend la $\Pr(\text{réussite de lecture})$ proche de 1.

Technique de lecture statistique (Couvét et al.)

- **A chaque période, la valeur d'horloge du site k est envoyée plusieurs fois au site i**
- **A partir des valeurs échangées et des délais d'aller-retour :**

- A l'aide des statistiques obtenues, le site i élabore un **modèle de l'horloge distante** du site k, avec deux paramètres a_{ik} et b_{ik} , ce qui permet d'avoir :

$$H_{ik}(t) = a_{ik}t + b_{ik}$$

- Le modèle élaboré permet de déterminer localement la valeur de l'horloge k sans demande quand le réseau est surchargé.

- **Le modèle de l'horloge k est raffiné grâce aux échanges de messages**
- **Exemples d'estimation de a_{ik} et b_{ik} :**

$$a_{ik} = \frac{\sum_{j=1}^n HS_{ij} \cdot HS_{kj} - n \cdot \overline{HS_i} \cdot \overline{HS_k}}{\sum_{j=1}^n (HS_{ij})^2 - n \cdot (\overline{HS_i})^2} \quad b_{ik} = HS_k(t_k) - a_{ik} \cdot (HS_i(t_i) - (D - \frac{d \min_{ki} - d \min_{ik}}{2}))$$

t_k instant de lecture de l'horloge k par le site k
 t_i instant de réception de HSK par le site i.

Calcul de la valeur de référence Selon les structures

- L'horloge du site **k** a été lue par le site **i** à l'instant $tlect_{ik}$
- En général, les instants de lecture sont différents : $tlect_{ij} \neq tlect_{ik} \quad (j \neq k)$
- Le site **i** calcule la référence à l'instant tr_i ($tr_i > tlect_{ik} \forall k$)
- L'horloge de **k** a évolué depuis sa lecture, il faut l'estimer à l'instant tr_i :

$$H_{ik}(tr_i) = H_{ik}(tlect_{ik}) + H_i(tr_i) - H_i(tlect_{ik})$$

Référence pour la structure décentralisée :

$$R_i(tr_i) = F(X_{i1}(tr_i), \dots, X_{in}(tr_i)) \quad X_{ik} \text{ correspond à } H_{ik} \text{ ou } HS_{ik}$$

$$\text{Condition : } \max\{|R_j(tr_j) - R_k(tr_k)| \mid 1 \leq j \leq n \quad 1 \leq k \leq n \quad j \neq k\} < \text{borne}$$

Référence pour la structure centralisée (p : site privilégié)

(1) $R_p(tr_p) = F(X_{p1}(tr_p), \dots, X_{pn}(tr_p))$

(2) $R_p(tr_p) = H_p(tr_p)$ Utilisation de l'horloge du site privilégié seulement

Tolérance aux fautes

- **Types de fautes**
 - Fautes temporelles (perte de messages, messages tardifs)
 - Fautes de valeurs dites non arbitraires (sites ou horloges défaillants)
 - Fautes arbitraires (fautes byzantines)
- **Fautes temporelles : détectées par l'utilisation de temporisateurs**
- **Fautes non arbitraires : utilisation de rejet**
 - **Rejet égocentrique :**
Si $|VH_{re\grave{c}ue} - ma_{VH}| > \text{Seuil}$ Alors Remplacer $VH_{re\grave{c}ue}$ par ma_{valeur}
 - **Rejet des valeurs aux extrémités :**
Trier les valeurs reçues, ensuite supprimer les m plus petites valeurs et les m plus grandes valeurs. m désigne le degré de tolérance aux fautes.
 - **Autres techniques**
- **Fautes arbitraires → utilisation de protocoles d'accord**

Technique de calcul global du moment d'ajustement Technique 1

- **Signature digitale**

- Signature de chaque message à l'émission.
- La signature est supposée infalsifiable.
- Avec la signature, les messages altérés sont rejetés.

- **Technique de Toueg et Srikanth**

- $j = 1, 2, \dots$ désigne le n° de tour de synchronisation. V désigne la période de synchronisation. T^0 = l'instant commun de démarrage de la synchronisation.
- Quand l'horloge d'un site arrive à l'instant $T^0 + j*V$, ce site diffuse un message signé « je suis prêt à ajuster mon horloge ».
- Chaque site relaie les messages correctement signés qu'il reçoit et il peut savoir s'il a déjà reçu un message en vérifiant sa signature. Si un message a déjà été reçu par un site, le message est rejeté, sinon ce site incrémente le compteur de messages reçus.
- Si le compteur de messages d'un site atteint la valeur m , ce site sait qu'il y a au moins un autre site qui est prêt à ajuster son horloge et ajuste son horloge par :
 $HS(t) = V*j + \beta$. (β désigne une marge d'erreur et m le degré de tolérance aux fautes)

Technique de calcul global du moment d'ajustement Autres techniques

- **Technique de diffusion centralisée**

- Périodiquement le site privilégié diffuse un message aux sites esclaves.
- A la réception du message du site privilégié, chaque site esclave ajuste son horloge.

- **Technique à base de synchronisation par le réseau**

- Certains réseaux (comme FIP) peuvent générer automatiquement et périodiquement un signal reçu par tous les sites du réseau.
- A la réception du signal du réseau, chaque site ajuste son horloge.

4. Fonction Ecart

- Rôle : calculer la donnée d'ajustement

Ecart pour la structure décentralisée (calculé par le site i) :

$$A_i(te_i) = R_i(tr_i) - HS_i(te_i)$$

te_i , moment de calcul
de la donnée d'ajustement

Ecart pour la structure totalement centralisée (calculé pour le site i) :

$$A_i(te_i) = R_p(tr_p) - X_{p,i}(tr_p)$$

p site privilégié
 $X_{p,i} = H_{p,i}$ ou $HS_{p,i}$

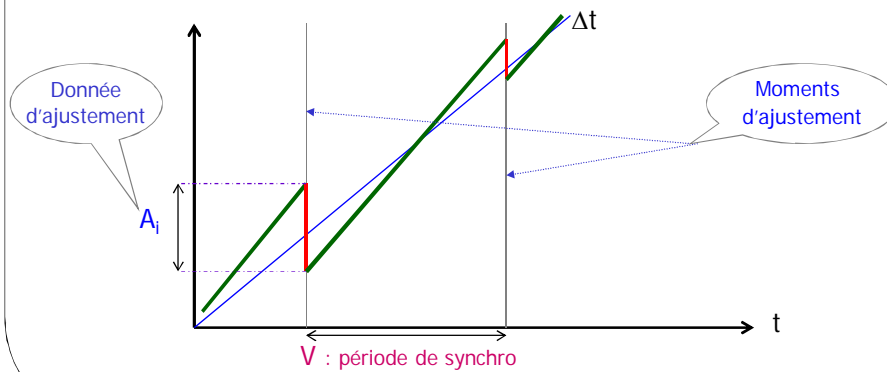
Ecart pour la structure partiellement centralisée (calculé par le site i) :

$$A_i(te_i) = R_p(tr_p) - X_i(te_i) + \alpha_i$$

$X_i = H_i$ ou HS_i
 α_i = délai moyen de communication
(en général)

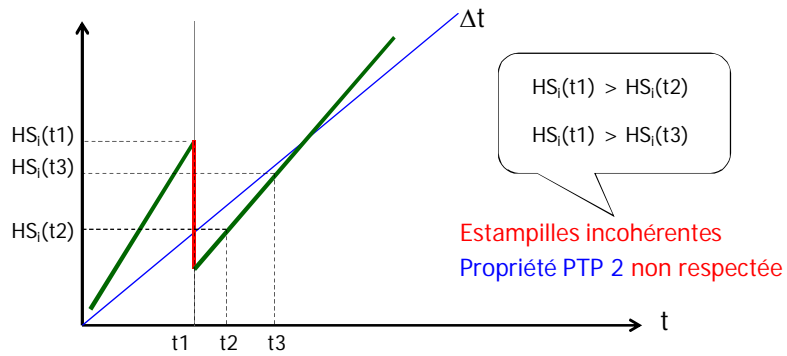
5. Ajustement d'horloge (1/3)

- Opération **locale** à chaque site
- Deux types d'ajustement : **discontinu** et **étendu** (pseudo-continu)



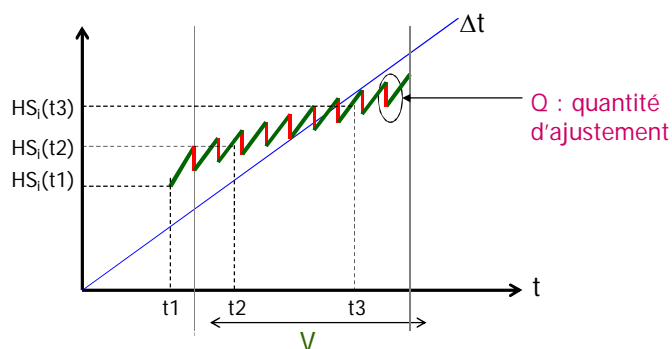
Ajustement d'horloge (2/3)

- Incohérence de datation provoquée la l'ajustement discontinu



Ajustement d'horloge (3/3)

- Ajustement étendu (pseudo-continu)



Forme 1 (il y a un nombre constant Z d'ajustements par période) $\Rightarrow Q = \frac{A(te_i)}{Z}$

Forme 2 (il y a un nombre variable d'ajustements par période) $\Rightarrow Q(t) = \frac{A(te_i)}{V} * (H_i(t) - H_i(ta_i))$

ta_i désigne l'instant du dernier ajustement

6. Analyse de performances

- **Considérer hypothèse par hypothèse, structure par structure,...**

- **Erreur maximale de lecture de l'horloge physique**

- Cas de la technique déterministe sans demande

$$ErrLectMAX = \frac{Gig}{2} + \rho \times D \max$$

- Cas de la technique déterministe avec demande

$$ErrLectMAX = \rho \times AR + \frac{AR}{2} - D \min$$

- Cas de la technique statistique de Couvet et al.

$$ErrLectMAX = \frac{(\rho + 1)AR}{2} - \frac{D \min_{ik} + D \min_{ki}}{2}$$

- **Précision d'accord**

- Cas de la structure décentralisée

$$PA = 2 \times \rho \times V + \max_{\substack{1 \leq i \leq n, \\ 1 \leq j \leq n, i \neq j}} (|R_i(tr_i) - R_j(tr_j)|)$$

- Cas de la structure totalement centralisée

$$PA = 2 \times ErrLectMAX + 2 \times \rho \times (2 \times S + V)$$

- Cas de la structure partiellement centralisée

$$PA = 2 \times ErrLectMAX + 2 \times \rho \times V$$

- **Exactitude de synchronisation**

$$EX = \rho$$

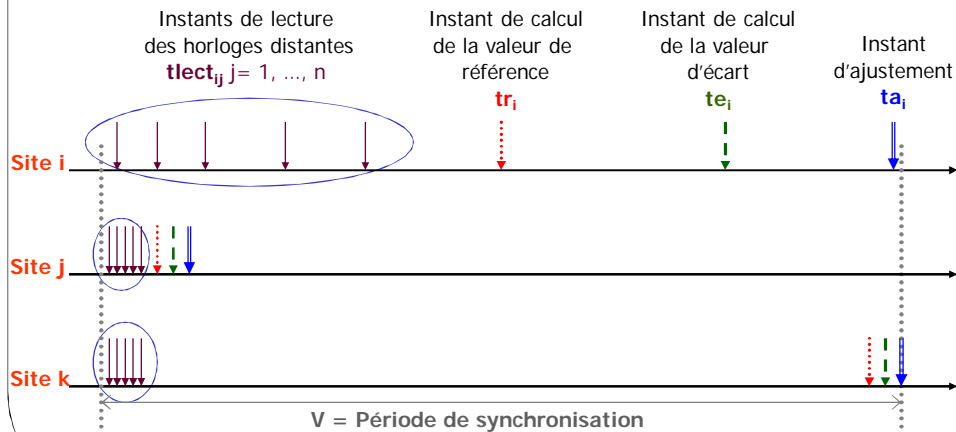
- **Précision d'accord minimale [Lynch et Lundelius 84]**

La limite théorique de la précision d'accord pour un système à n sites est :

$$PA \geq (D \max - D \min) \left(1 - \frac{1}{n}\right)$$

Précautions d'implantation

- Limiter le temps pris par l'exécution de l'algorithme de synchronisation



7. Conclusion

- La synchronisation d'horloges = **Problème important et complexe**
- La synchro d'horloges est nécessaire dès le moment où des instants absolus sont utilisés par les sites pour coordonner leurs activités. Elle n'est pas nécessaire si les sites ont des horloges physiques correctes et elles n'utilisent que des durées (timers).
- Différentes formes d'algorithmes
- Difficultés de déterminer avec précision les performances des algorithmes à cause de la variété des hypothèses
- Sélection automatique des algorithmes en fonction des performances requises (précision d'accord, complexité...)