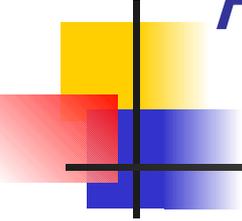


# Priorité et fiabilité dans le système MoSAIC

Damien Martin-Guillerez

IRISA - ACES

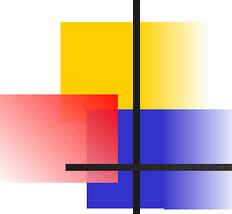
*ACI MoSAIC (LAAS – INRIA - Eurecom)*



# Approche

---

- Quels sont les types de données générées par ces terminaux ?
- Quels sont les caractéristiques pertinentes de ces données ?
- Quels sont les points difficiles ?



# Caractéristiques des données

---

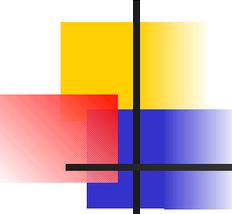
- La présence de dépendances
  - Dépendances « temporelles » : les nouveaux blocs dépendent des anciens.
  - Dépendances « spatiales » : les différents blocs sauvegardés à un moment dépendent l'un de l'autre.
- La priorité
  - Fonction de la méthode de production et de l'importance de la donnée
- Les modifications réalisables
  - Création (C) (toujours faisable)
  - Modification (W)
  - Suppression (D) (toujours faisable)
  - Ajout d'un delta (U)
- La taille
- Donnée dont la durée de vie est limitée

# Les type de données

Type	Taille	Dépendances	Priorités	W	U	
SMS	<200o	Fil de discussion	Moyenne			
MMS	~50ko	Fil de discussion	Moyenne			
Mail	1-5ko <10Mo	Fil de discussion	Faible (sauf draft)			
Tasks	<200o	Possibles	Forte	X		Durée de vie limitée
RDV	<1ko. DB~100ko	Non nécessaires	Moyenne	X	X	
Vidéos	~10Mo <700Mo	Aucune	?			
Photos	~50ko <2Mo	Aucune	?			
Audio	~3Mo	Aucune	?			
Notes	<10ko	Possibles	Forte	X	X	

# Problématique

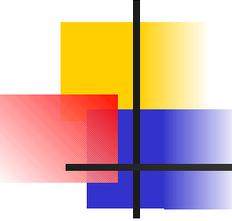
---



# Influence de la taille

---

- Il y a une limite sur la taille des données :
  - Pour que les transmissions se passent bien
    - Débit
    - Temps de connectivité [Troel2004]
  - Analyse statistique pour calculer la taille maximale des paquets ?
- Pour les données de grande taille :
  - Delta compression => dépendances « temporelles »
  - Fragmentation => dépendances « spatiales »
  - L'utilisation de l'algorithme de dispersion de Rabin [IDA] permet de fragmenter en ajoutant de la redondance



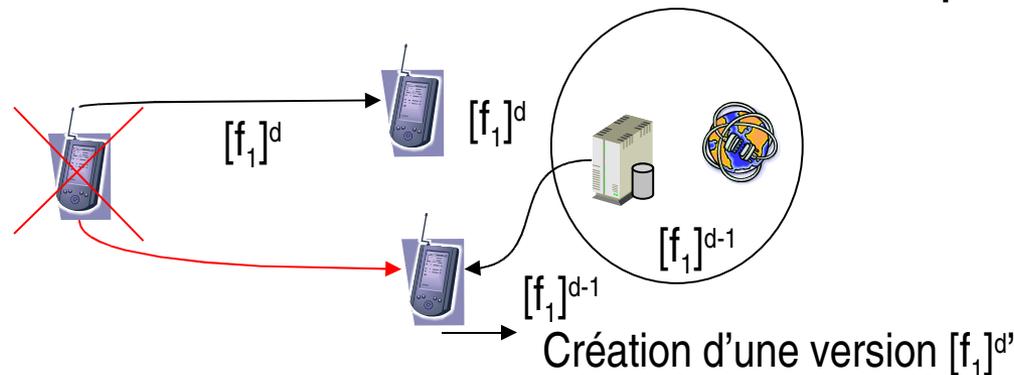
# Durée de vie des sauvegardes

---

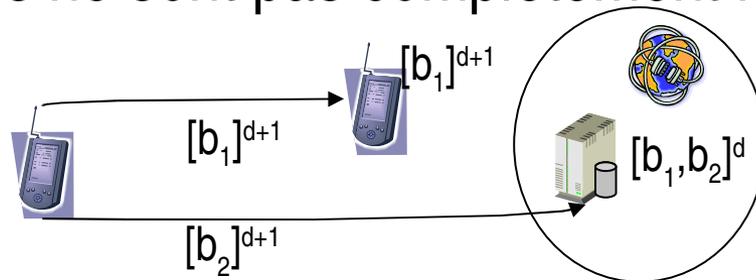
- Durée de vie des sauvegardes variable
  - Dépend des données...
  - ... et des circonstances.
    - Arrivée d'une nouvelle version [EFS]
    - Suppression pour libérer de l'espace (Méthodes de caching)
    - Suppression de la sauvegarde en fonction du nombre de duplicata
    - Etc...

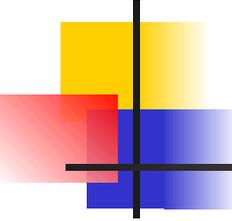
# Cohérence et restitution

- Résolution des conflits dus aux défaillances par [HotSync, Bayou]



- Nécessité de conservation d'anciennes versions lorsque les dépendances ne sont pas complètement résolue



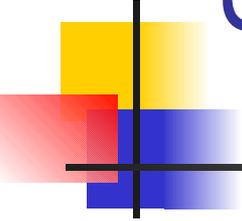


# Fiabilité d'une sauvegarde

---

- Estimation de la fiabilité d'une sauvegarde
  - Pour estimer le niveau de duplication nécessaire
  - Pour libérer de l'espace de sauvegarde (caching)
  
- Et de la priorité des données
  - Pour le niveau de duplication...
  - Pour choisir les données à sauvegarder
  - Elle dépend des dépendances...

# Un algorithme de sauvegarde



---

- Chaque donnée est associée à une certaine fiabilité.
- Chaque sauvegarde augmente la fiabilité atteinte.
- La différence entre la fiabilité atteinte et celle demandée donne la priorité de chaque donnée pour être sauvegardée.
- Dans le cas de plusieurs coopérateurs, on choisit celui ayant la meilleure probabilité de réussite.

# Modification de la fiabilité des sauvegardes

- Découpage en  $k$  paquets dont  $n$  nécessaires pour reconstruire le fichier ( $k \geq n \geq 1$ ).

- Si on sauvegarde un  $k^{\text{ème}}$  paquet :

- $P_k^n = (1 - P_k) \cdot P_{k-1}^n + P_k \cdot P_{k-1}^{n-1}$

- $P_k$  est la fiabilité du  $k^{\text{ème}}$  paquet (fiabilité du terminal...)

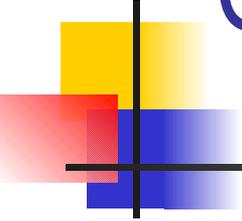
- En particulier :

- $P_n^n = \prod_{i=1}^n P_i$

- $P_k^n = 0$  si  $k < n$  et  $P_k^0 = 1$

# Fiabilité des sauvegardes

- Ceci est valable si chaque sauvegarde est indépendante.
  - Cas sur des terminaux différents
  - Sinon la probabilité est plus élevée => on a une minoration
- Pour les dépendances temporelles, la fiabilité de la sauvegarde est le produit de celles de la sauvegarde précédente et du delta envoyé.

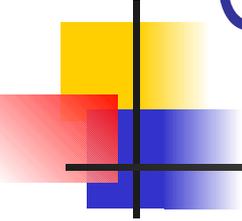


# Ce qu'il faut maintenant...

---

- Estimer la probabilité de chaque terminal (mesure donnée par l'incitation à la collaboration ?)
- Intégrer l'utilisation des ressources (énergie et espace mémoire).
- Intégrer les suppressions pour libérer de l'espace de sauvegarde (calcul de la probabilité ?).

# Questions ?



---

*Three things are certain:  
Death, taxes, and lost data.  
Guess which has occurred.*