Sauvegarde coopérative entre pairs pour dispositifs mobiles

Ludovic Courtès,
Marc-Olivier Killijian, David Powell, Matthieu Roy

Contexte

Le projet MoSAIC

- Projet de **3 ans** démarré en sept. 2004 : IRISA, Eurecom et LAAS-CNRS
- Partiellement financé par l'**ACI S&I 2004**

Cibles

- **Dispositifs mobiles communicants** (PDA, téléphones, etc.)
- **Réseaux mobiles ad hoc**, spontanés ; interactions de type **pair à pair**

Objectifs de sûreté de fonctionnement

- Amélioration de la **disponibilité des données**
- Garantie de leur **confidentialité** et **intégrité**

- **Introduction**
  - Motivations
  - Exemple de scénario
  - Les défis majeurs
- Objectifs et problématique
- L'apport du pair-à-pair
- Conclusions et pistes de recherche

Motivations

Sauvegarde compliquée et coûteuse

- Sur sa **machine de bureau**
- À distance : accès à Internet **intermittent**, machine de bureau **parfois inaccessible**, **communications coûteuses** (GPRS, UMTS), etc.

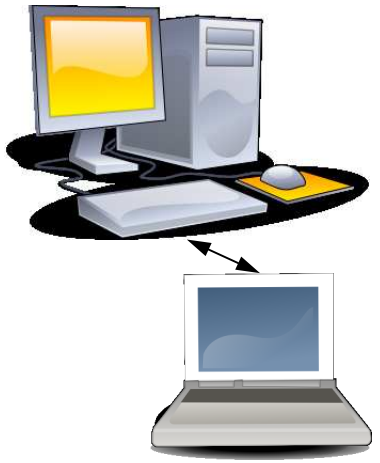
Notre approche : la sauvegarde *coopérative*

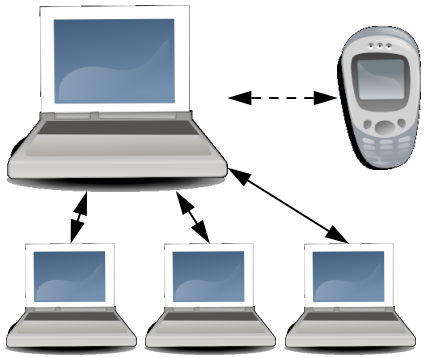
- Tirer parti des rencontres **de manière opportuniste**
- **Débit** et **faible coût** (énergétique) des communications courtes distances (Wifi, Bluetooth, etc.)
- Profiter des **ressources excédentaires disponibles**
- **Diversité des modes de défaillance**
- Ensemble **« autogéré »**

⇒ **De « pair-à-pair » à « face-à-face » (Kortuem et al.)**

Exemple de scénario

1. Avant de partir : **synchronisation du mobile**
2. Pendant le voyage : **modification/création de documents sur le mobile**
3. Pendant le voyage : **rencontres et échange de données nouvelles**
4. Avant de rentrer, **le drame : perte du mobile**
5. Au retour : **récupération des données stockées de manière coopérative** !





Les défis majeurs

- **Accès intermittent** à une infrastructure
- **Interactions éphémères**
- Énergie, puissance et stockage **limités**
- Pas d'**organisation** préalable
- **Aucune relation de confiance** a priori

- Introduction
- **Objectifs et problématique**
  - Les fonctions du service
  - Sûreté de fonctionnement du service
- L'apport du pair-à-pair
- Conclusions et pistes de recherche

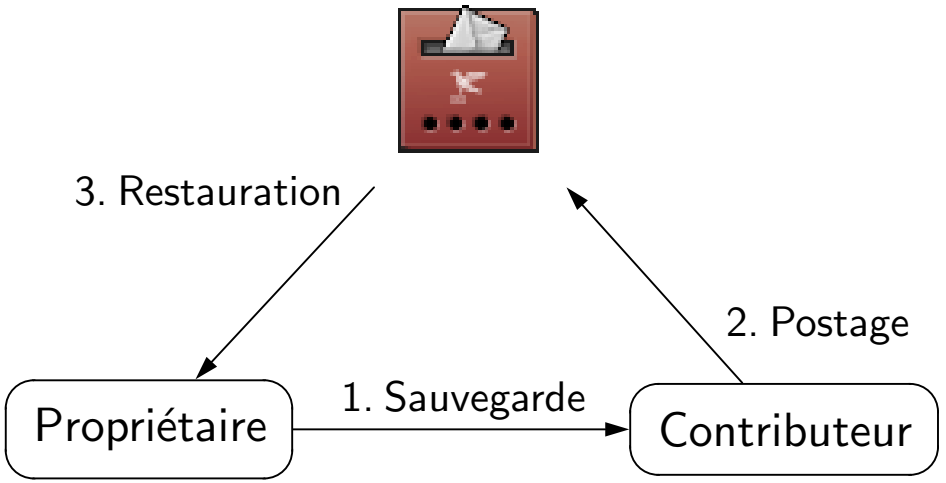
Les fonctions du service

Découverte et allocation de ressources

- Tenir compte du **coût de la communication**
- ... de la **densité de nœuds mobiles**
- ... de la **volatilité des connexions**

Recouvrement des données

- L'approche **pro-active** : utilisation d'une « **boîte aux lettres** »
- L'approche **réactive** : le propriétaire **interroge le réseau**



Sûreté de fonctionnement du service

Intégrité et cohérence des données sauvegardées

- Détection des **corruptions accidentelles**
- Détection des **corruptions intentionnelles**

Confidentialité

- Recours à la **fragmentation** et **dissémination** des données
- Permet l'utilisation de **méthodes de chiffrement peu coûteuses** en énergie

Disponibilité

- **Duplication** des données chez plusieurs contributeurs
- Résistance au **déni de service** : égoïsme, inondation, rétention
- Choix des **données à effacer**

- Introduction
- Objectifs et problématique
- **L'apport du pair-à-pair**
  - S'inspirer de l'état de l'art
  - Découverte et allocation de ressources
  - Réduction de la duplication inutile des données
  - Intégrité des données
  - Garantie de la confidentialité
  - Résistance aux attaques en déni de service
- Conclusions et pistes de recherche

S'inspirer de l'état de l'art

Systèmes de sauvegarde pair à pair

- *Pastiche* et *PeerStore* (sur Internet), *FlashBack* (pour un PAN)
- **Incitations à la coopération**, établissement de la **confiance**, etc.

Partage de fichiers pair à pair

- *GNUnet*, *Freenet*, *OceanStore*, etc.
- **réseaux virtuels** : DHT, non structurés

Gestion de versions et archivage

- *Venti*, *Elephant File System*, etc.
- **Indexation**, **gestion des ressources** sur le long terme

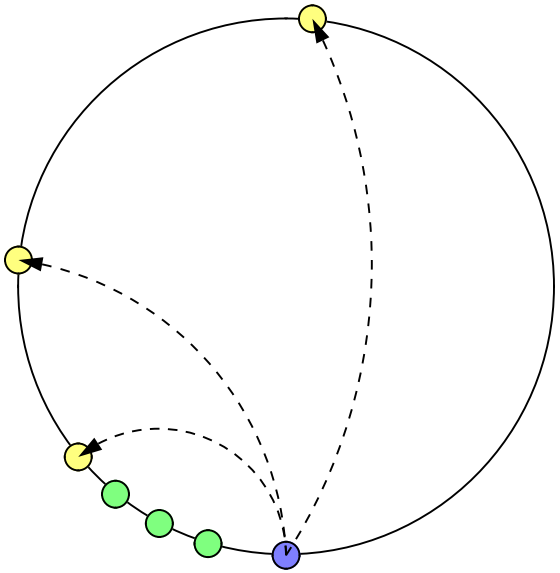
Découverte et allocation de ressources

1ère approche : groupes de partenaires

- Stockage réparti **au sein du groupe**, statique
- Pratique sur **Internet** ou dans un **PAN** (voisinage constant)

2ème approche : réparti entre tous les participants

- **Réseau virtuel**, **DHT**
- Fait l'**hypothèse d'un coût de communication faible**
- Hypothèse d'un **ensemble de participants assez fermé**

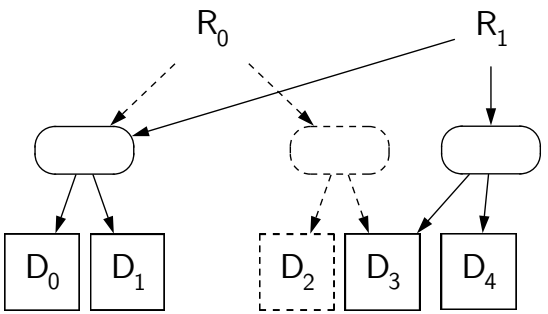


Réseau virtuel dans *Chord*

Réduction de la duplication inutile des données

Stockage indexé par son contenu

- bloc de données $\rightarrow$ **identifiant fonction de son contenu** (e.g. SHA1)
- Deux blocs identiques ont le **même identifiant**
- Donc chaque bloc n'est **stocké qu'une seule fois**
- **Économie d'énergie** : échange des *nouveaux* blocs



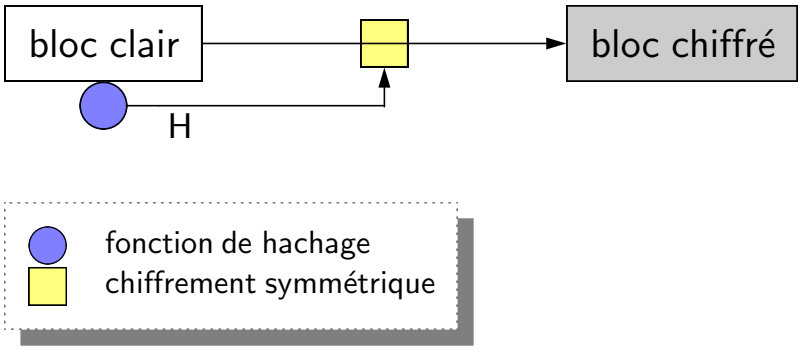
Découpage en blocs et encodage

- Création d'un **arbre d'identifiants** de blocs (*Merkle tree*)
- L'**identifiant du bloc racine** suffit pour adresser l'ensemble des blocs

Intégrité des données

- Stockage indexé par le contenu $\rightarrow$ **vérification directe des blocs** (comparaison identifiant vs. condensé)
- Encodage en arbre $\rightarrow$ **vérification d'intégrité incrémentale** (bloc par bloc)
- **Déchiffrement $\Rightarrow$ vérification d'intégrité**

Garantie de la confidentialité



Comment chiffrer et garantir le stockage à instance unique ?

- Recours au ***chiffrement convergent***
- chiffrement symétrique des blocs avec une **clef fonction de leur contenu** (un condensé)
- pour déchiffrer un bloc, **il faut connaître (avoir connu) son contenu (son condensé)**

⇒ **Stockage à instance unique *chiffré***

⇒ **Mais affaiblissement de la confidentialité**

Résistance aux attaques en déni de service

Égoïsme et inondation

- **Coût partagé entre tous** (cas des DHT) : problème de la **« tragédie des communaux »**
- **Échanges symétriques** : **ne passe pas à l'échelle** (grand nombre de rencontres)
- **Technique du « troc »** (échange d'obligations) : **valable pour un monde clos**
- Mécanisme de **confiance, réputation, ou micro-économie** (partage de fichiers entre pairs) : piste à privilégier

Rétention de données

- **Tolérer** les déconnexions non intentionnelles, **punir** les abus
- Lancer régulièrement des **défis**
- **Punition progressive** des indisponibilités

- Introduction
- Objectifs et problématique
- L'apport du pair-à-pair
- **Conclusions et pistes de recherche**
  - Limites du parallèle avec le pair-à-pair
  - Adaptation aux moyens de communication
  - Modèle de confiance

Limites du parallèle avec le pair-à-pair

Limites

- **Qualité et bande passante** des connexions
- **Évolution constante de l'entourage** d'un dispositif
- **Absence de mécanisme de désignation fixe** (IP, etc.)
- **Consommation énergétique** $\rightarrow$ **adapter les protocoles**

Spécificités : la découverte de ressources

- Sur Internet : utilisation de **listes de participants**, ou **diffusion locale**
- Environnement sans fil : **découverte de participants physiquement proches**
- S'inspirer d'algos de **routage ad hoc** (ressource = table de routage)
- S'inspirer de **systèmes de fichiers répartis en mode ad hoc**

Adaptation aux moyens de communication

- **Profiter des accès intermittents à Internet**
  - Déchargement vers les « **boîtes aux lettres** »
  - **Synchronisation** avec machine de bureau, si possible
- **Sur Internet** (mécanisme de boîte aux lettres)
  - Approche « classique » **pair-à-pair**
  - Ou approche **centralisée** (serveur de sauvegarde)

Modèle de confiance

Les paramètres importants

- **décentralisation** intrinsèque
- **voisinage hautement dynamique**
- **nombre élevé de partenaires**

$\Rightarrow$ Mécanisme de réputation auto-portée ?

Cohabitation avec Internet (boîtes aux lettres)

- Comment **récompenser les machines de bureau ?**

Fin.

Questions ?

http://www.laas.fr/mosaic/