Multidistribution Multi-sources Adaptative

Bertrand Ducourthial
UMR CNRS 6599 HEUDIASYC
Université de Technologie de Compiègne
Bertrand.Ducourthial@utc.fr

Sébastien Tixeuil
UMR CNRS 8623 LRI
Université de Paris-Sud
tixeuil@lri.fr

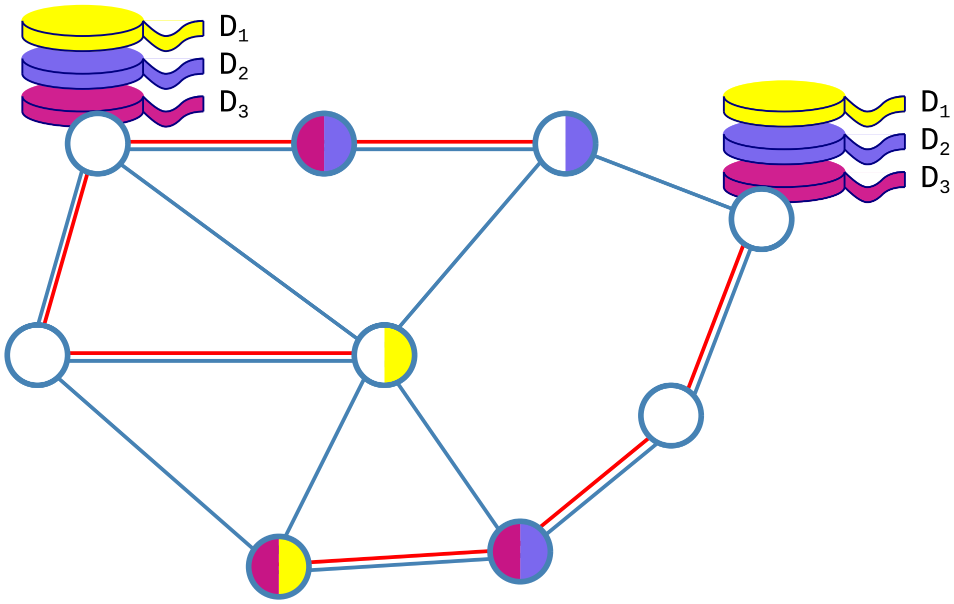
25 mai 2001

Résumé

En utilisant un modèle générique d'algorithmes tolérant les défaillances, nous construisons un schéma de distribution sélective de données. L'algorithme optimise la distribution en fonction d'un critère donné, gère plusieurs groupes de destinataires et plusieurs sources, tolère les défaillances temporaires et s'adapte dynamiquement aux évolutions du réseau.

UMR CNRS 6599 Heudiasyc, Université de Technologie de Compiègne

Multidistribution multi-source (1/2)



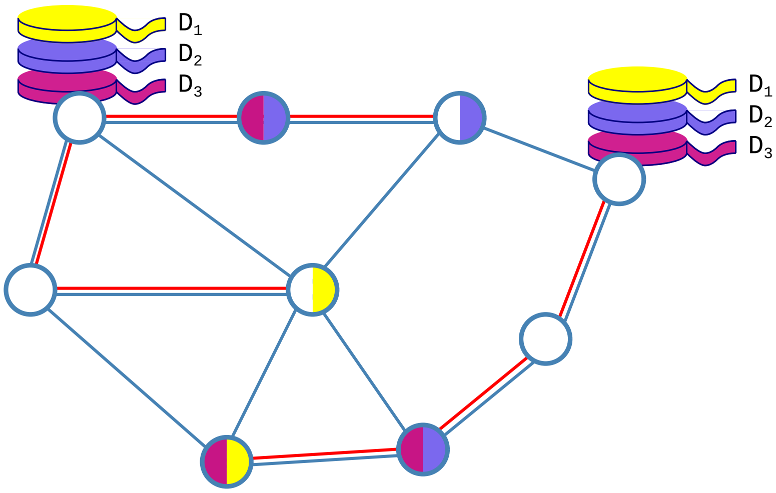
Contexte

- Serveurs ftp, vidéo à la demande, sites mirroirs...
- Réseaux filaires ou non

Multidistribution multi-source

- plusieurs sources, même information sur les sources
- plusieurs unités de données $D_1 \dots D_k$
- à chaque D_i correspond un groupe de destinataires G_i, éventuellement vide
- un site est dans 0, 1 ou plusieurs groupes

Forêt de multidistribution (2/2)



- But
  - choix de la meilleure source et du meilleur chemin
  - adaptation aux variations (infrastructure, trafic...)
  - utilisation de métriques variées
  - minimisation des diffusions sur la forêt
    $D_i \cap D_j$ diffusé une seule fois aux sites $G_i \cap G_j$
  - groupes dynamiques
  - tolérance des défaillances transitoires
  - pertes équitables, duplications finies et déséquencements arbitraires sur le lien supportées (messages de contrôle)
  - messages de contrôle non acquittés
- Moyens
  - auto-stabilisation
  - algorithme générique paramétré, r-opérateurs

Auto-stabilisation : définition

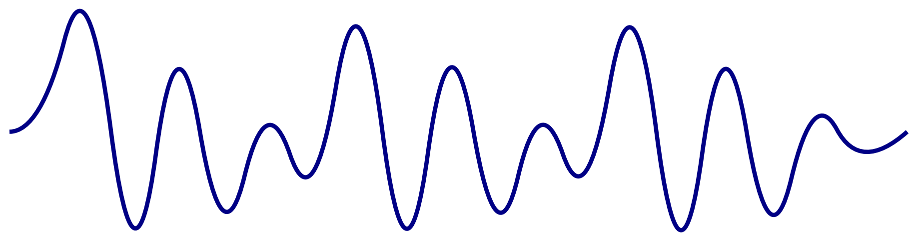
- Fautes temporaires
  - faute $\rightsquigarrow$ modification des données
  - registres, RAM susceptibles d'être corrompues
  - ROM et code non altérables

  $\rightsquigarrow$ modèle général pour supporter les modifications des taux (évolution du trafic, du réseau, panne...)
- Rétablissement après panne
  - une fois que les fautes cessent, le système retrouve un comportement correct en un temps fini [Dijkstra 74, Dolev 00]
  - initialisation quelconque

Auto-stabilisation : limitation

- Temps de convergence
  - valeur erronée sur un nœud
    $\leadsto$ tous ses descendants peuvent être affectés
  - si absence de circuits, $O(D)$
  - topologie quelconque $O(D + |\mathbb{S}|)$

- Inconvénient
  - forêt mouvante



- Parade : espacer les variations
  - modification du réseau, pannes $\leadsto$ ok
  - évolution du trafic
    $\leadsto$ moyennes mobiles, seuil d'alerte

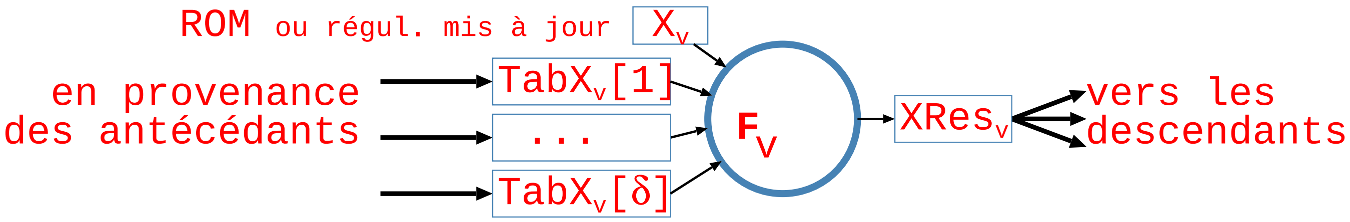
État de l'art

- Auto-stabilisation
  - graphes non orientés [Dolev *et al.* 93, Afek *et al.* 90]
  - graphes orientés fortement connexe [Alstein *et al.* 94]

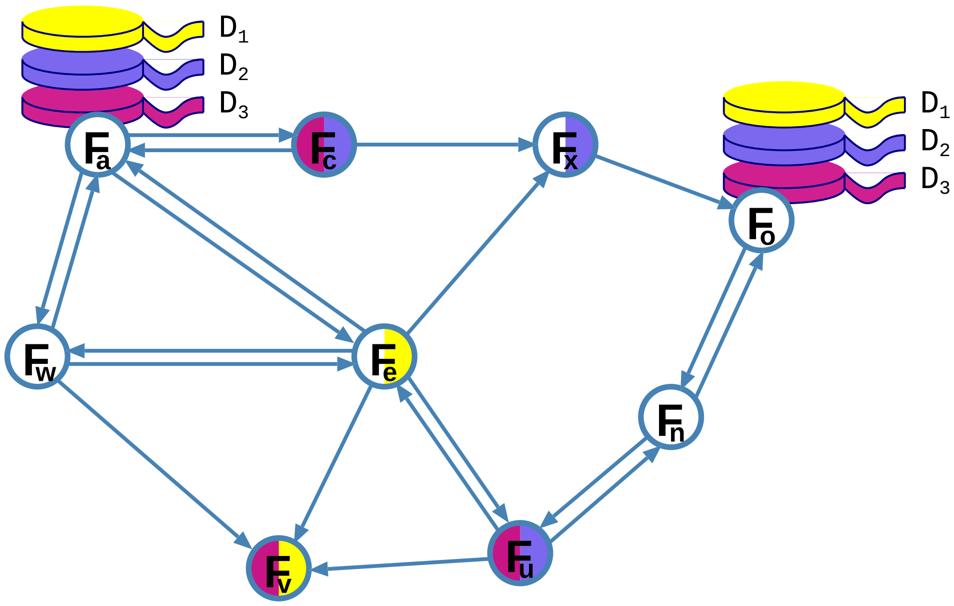
    $\leadsto$ topologie virtuelle connue ($e.g.$ arbre)
- Tâches silencieuses
  - convergence vers une configuration et non une exécution
  - existence d'une configuration terminale [Dolev *et al.* 96]
- Tâches silencieuses auto-stabilisantes
  - graphes orientés, non fortement connexe
  - mais sans circuit $\leadsto$ DAG [Das *et al.* 99]
- Notre travail
  - tâches silencieuses auto-stabilisantes sur tout graphe orienté
  - s'il existe un opérateur (ou une séquence d'opérateurs) spécifique pour instancier l'algorithme

Algorithme paramétré

- Algorithme



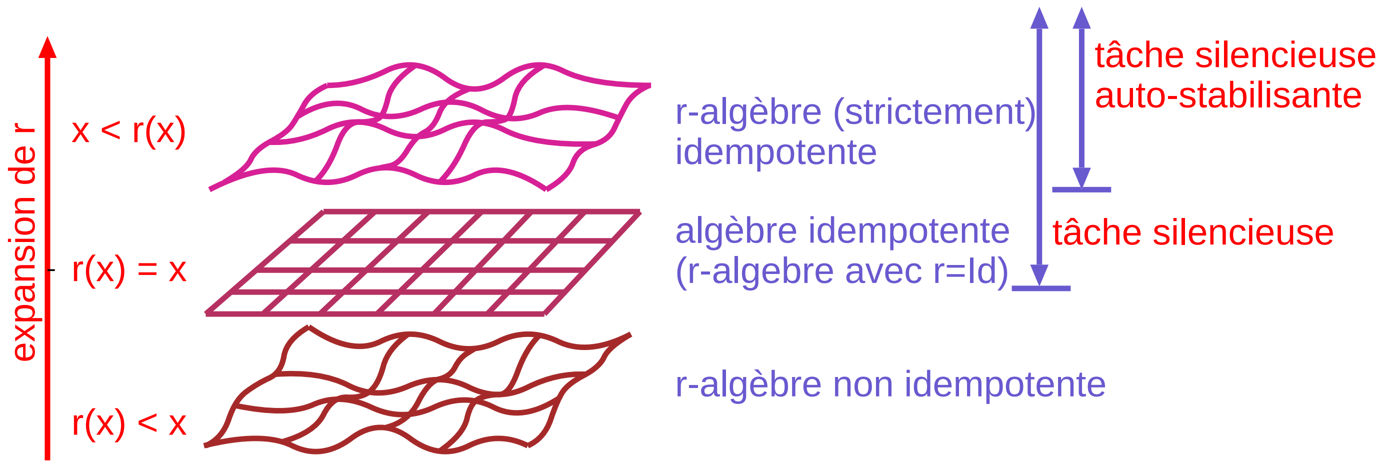
$$\texttt{XRes}_v \leftarrow \mathbf{F}\left(\texttt{X}_v, \texttt{TabX}_v[1], \ldots, \texttt{TabX}_v[\delta^-]\right)$$



- Paramètres ($i.e.$ opérateurs F) ?
  - tâche silencieuse ($i.e$ convergence vers une configuration)
  - auto-stabilisante ($i.e.$ convergence en dépit des pannes)
  - conforme aux spécifications ($e.g.$ forêt meilleurs taux)

r-opérateurs

- Infimum (s-opérateurs)
  - infimum $\rightsquigarrow$ tâches silencieuses [Tel91]
    ($e.g.$ min(x, y))
  - infimum $\Rightarrow$ non auto-stabilisant [OPODIS98, DC]
- r-opérateurs
  - généralisation des infimums [SIROCCO98]
    ($e.g.$ minc$(x, y) =$ min$(x, y + 1)$)
  - r-opérateurs strict. idempotent $\rightsquigarrow$ auto-stabilisation
    atomicité read-write si ordre total [OPODIS98, DC]
    ordre partiel si démon total. réparti [SIROCCO00, TCS]



Conditions sur l'opérateur

$$\min(x_0, x_1 + d_1, \ldots, x_n + d_n)$$

- Réseau sans circuit
  $\rightsquigarrow$ tout r-opérateur

  1. $\mathbb{F} : \mathbb{S}^n \mapsto \mathbb{S}$
     $\mathbf{F}(x_0, \ldots, x_n) = x_0 \oplus r_1(x_1) \oplus \cdots \oplus r_n(x_n)$
  2. $\mathbb{S}$ est fini
  3. $\oplus$: infimum $\rightsquigarrow$ ordre total $\preceq_\oplus$
  4. $r_i(x \oplus y) = r_i(x) \oplus r_i(y)$ (homomorphismes)

- Réseau avec circuit
  $\rightsquigarrow$ tout r-opérateur idempotent
  5. $x \prec_\oplus r_i(x)$

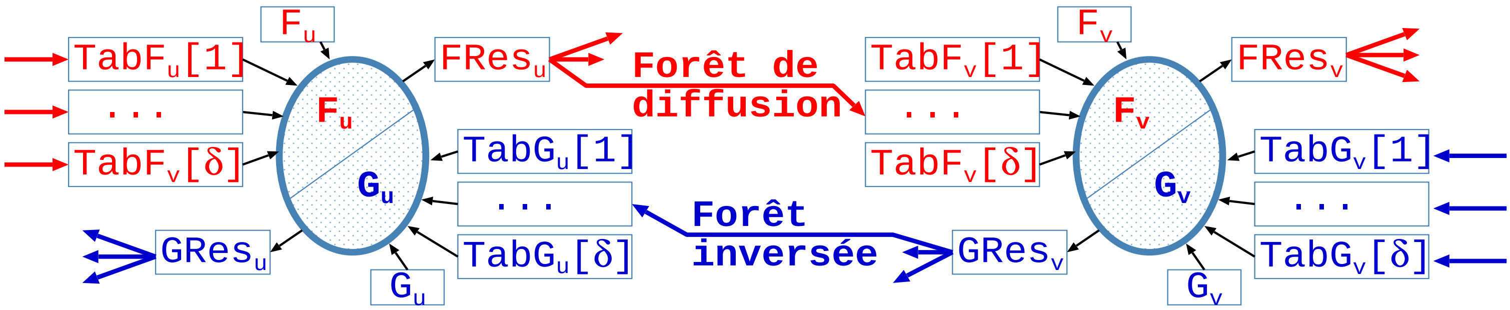
- Résultat du nœud v

$$\bigoplus \{r_{P_{u\to v}}(\mathtt{X}_u),\ u \text{ antécédant de } v \text{ dans le réseau}\}$$

$$\min \left\{ \mathtt{X}_u + \sum_{(p,q)\in P_{u\to v}} d_{(p,q)},\ u \text{ antécédant de } v \right\}$$

Multidistribution multi-sources

- **1. Construction de la foret**
  - Forêt couvrant l'ensemble du réseau
    - ⇝ Possibilité de circuit
    - ⇝ r-opérateurs strictement idempotents
  - Métrique ⇝ poids sur les arcs
  - *Multi-source shortest path problem*
- **2. Construction des groupes**
  - Sur la forêt inversée
    - ⇝ pas de circuit
    - ⇝ r-opérateurs idempotents

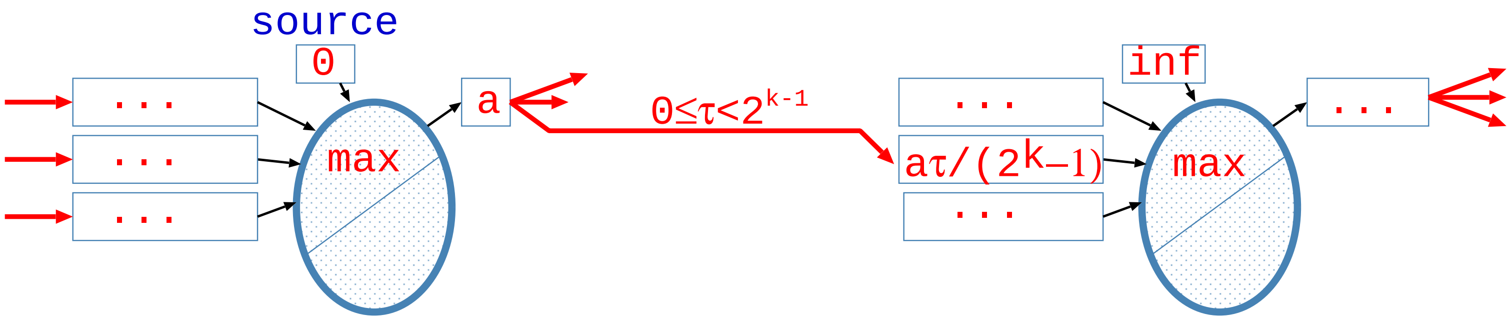


Construction de la forêt

- **Métrique**
  distance, nombre de sauts, débit,
  taux de perte de paquets...

- **Exemple : taux** [Algotel00]
  - variables de k bits $\leadsto \mathbb{S} = \{0, \dots, 2^k - 1\}$
  - $0 \leq \tau < 2^k - 1$
    ligne hors-service ligne parfaite

- **Opérateur**
  - taux d'un chemin : produit $\leadsto r(x) = x \times \tau / (2^k - 1)$
  - meilleur chemin : maximum $\leadsto \oplus \equiv \max$

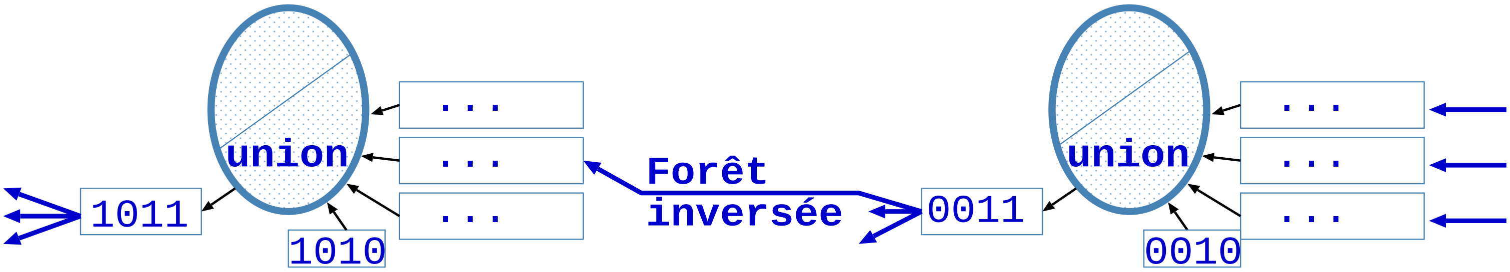


- **Conditions vérifiées**
  - homomorphisme $\max(x, y) \times \tau = \max(x \times \tau, y \times \tau)$
  - stricte idempotence $x \prec_{\max} r(x)$ $i.e.$ $x > r(x)$

Construction des groupes

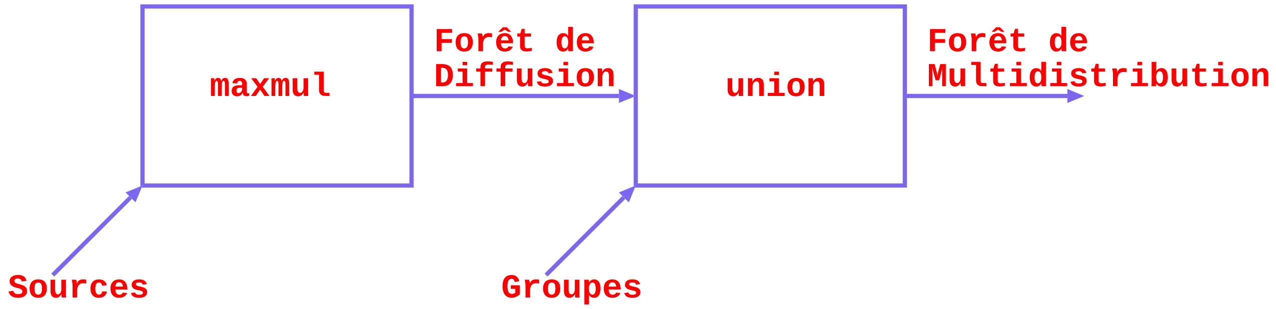
- Appartenance du site u au groupe G_i
  - i$^{\text{ème}}$ bit de $\mathtt{G}_u$ à 1.
  - variables de l bits pour l groupes

- Groupes à diffuser dans une branche de la forêt
  - union des groupes auxquels appartiennent les sites de la branche
  - union $\equiv$ r-opérateur idempotent (s-opérateur)
  - union $\equiv$ ou bit-à-bit des variables $\mathtt{G}_u$, $u \in$ branche
  - non strictement idempotent, mais pas de circuit dans la forêt inversée.



Fonctionnement de la multidistribution

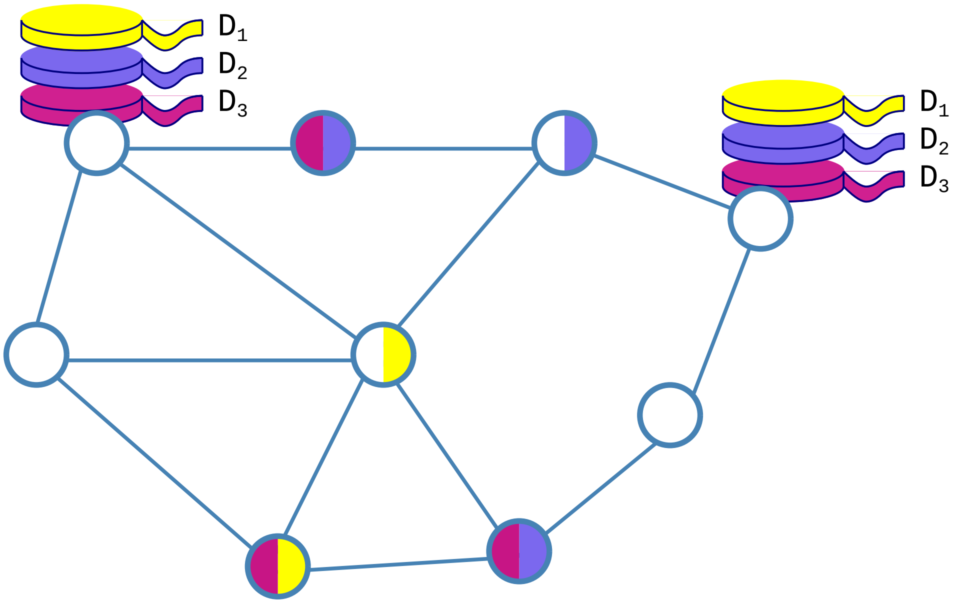
- Algorithme
  - Construction de la forêt
  - Construction des groupes
  - Distribution des données



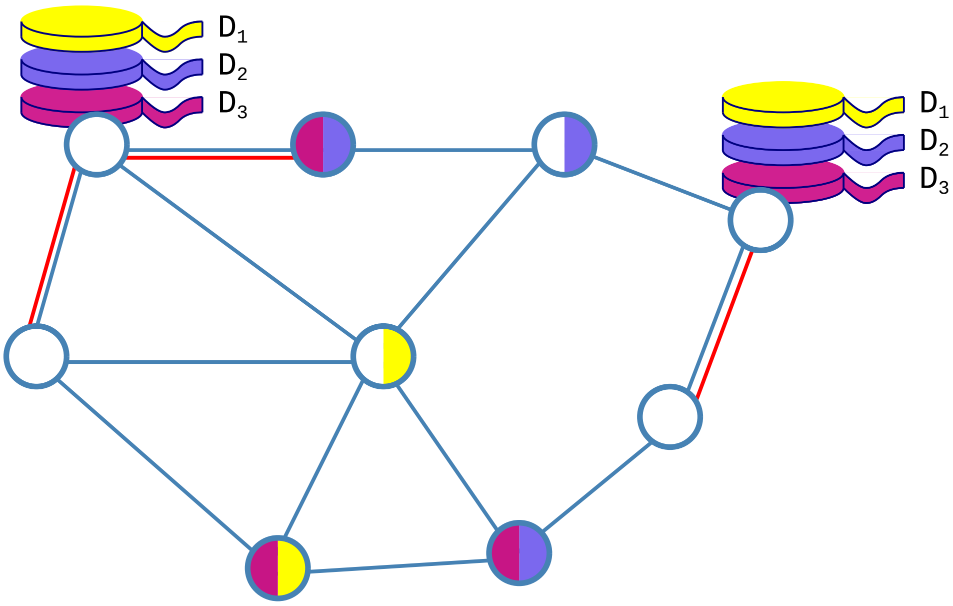
- Distribution des données
  - messages $\langle \mathtt{G}, \mathtt{M} \rangle$
  - réception par le site u
    - si $\mathtt{G}_u \cap \mathtt{G} \neq \emptyset$ alors
      $\mathtt{M}$ délivré à la couche applicative de u
    - pour chaque fils v de u dans la forêt,
      si $\mathtt{TabG}_u[\mathrm{num}(v)] \cap \mathtt{G} \neq \emptyset$ alors
      envoyer $\langle \mathtt{G} \cap \mathtt{TabG}_u[\mathrm{num}(v)], \mathtt{M} \rangle$ à v.

Exemple : forêt (1/2)

- Initialement

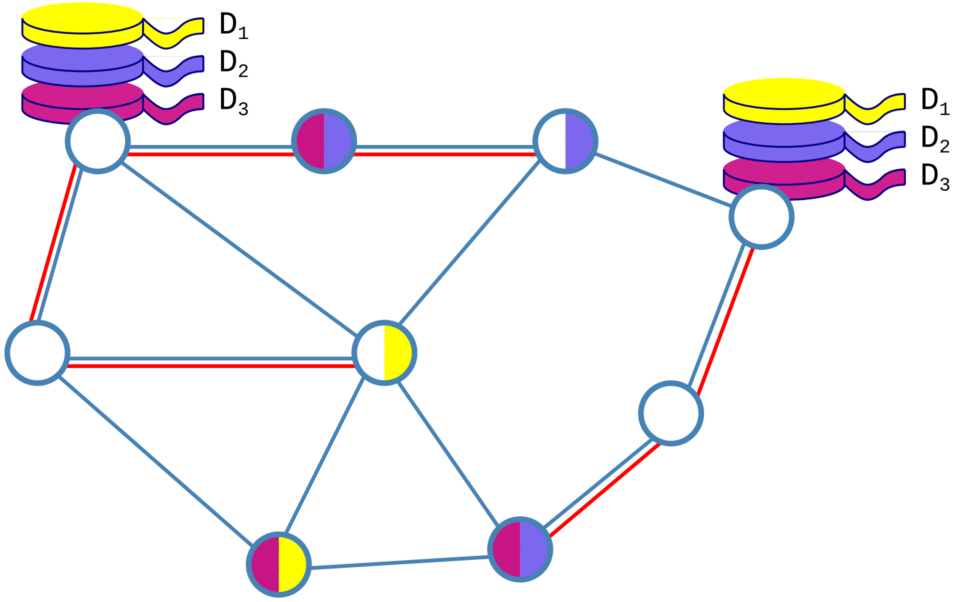


- Après une étape

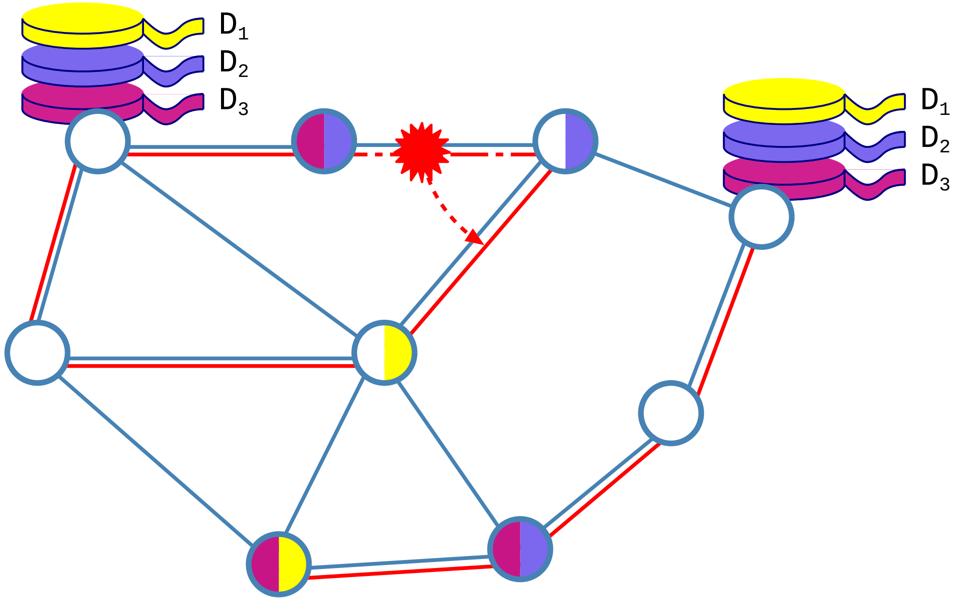


Exemple : forêt (2/2)

- Après deux étapes

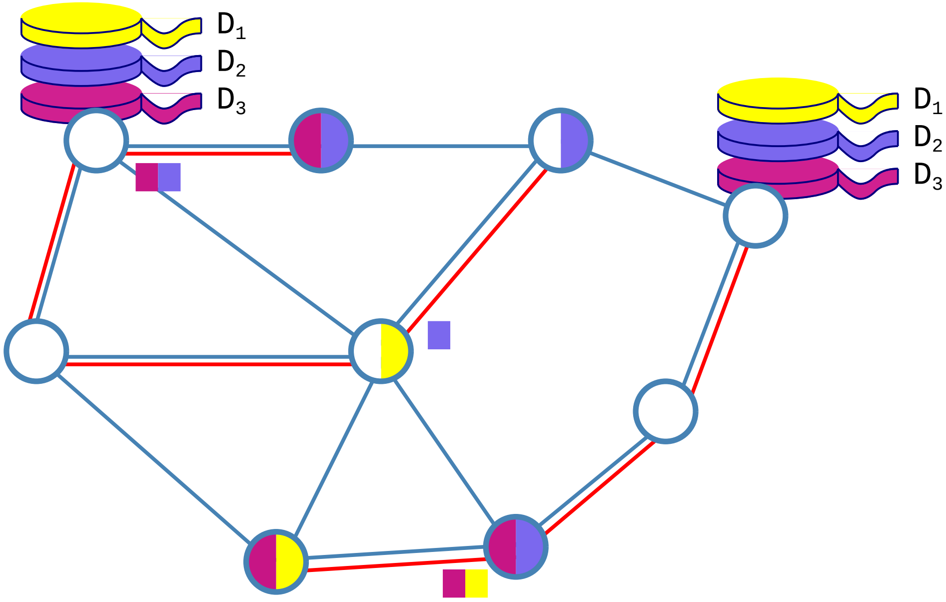


- Adaptativité

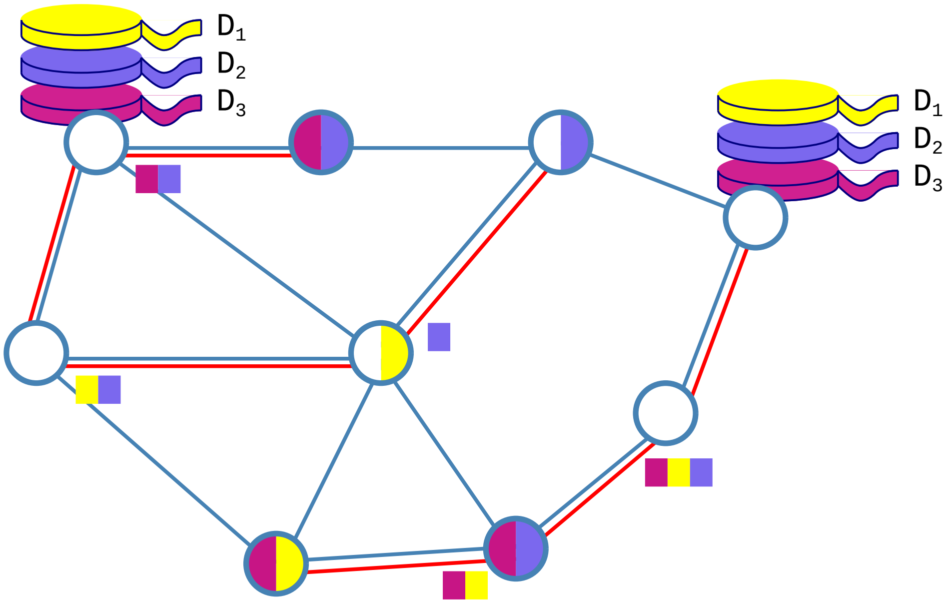


Exemple : groupes (1/2)

- Après une étape

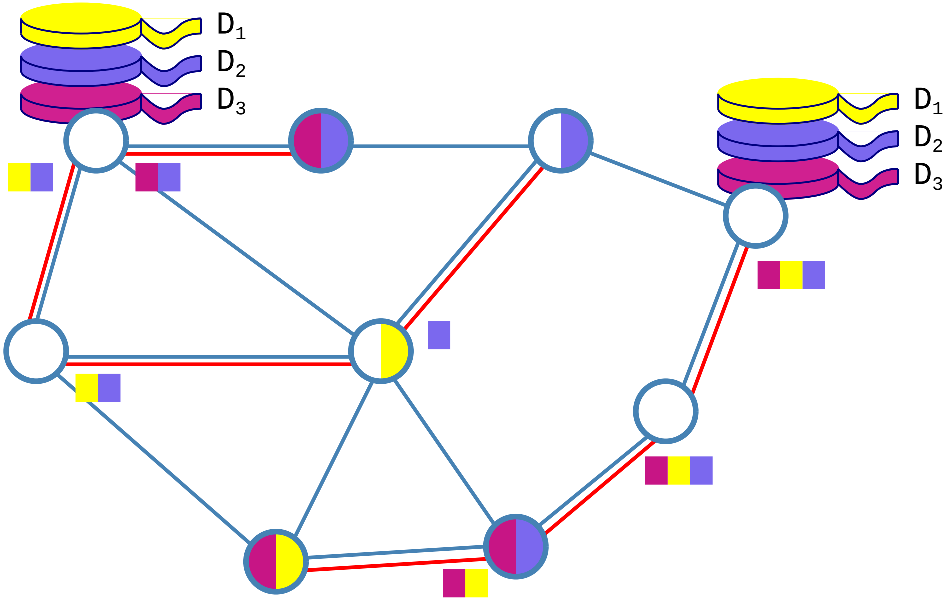


- Après deux étapes

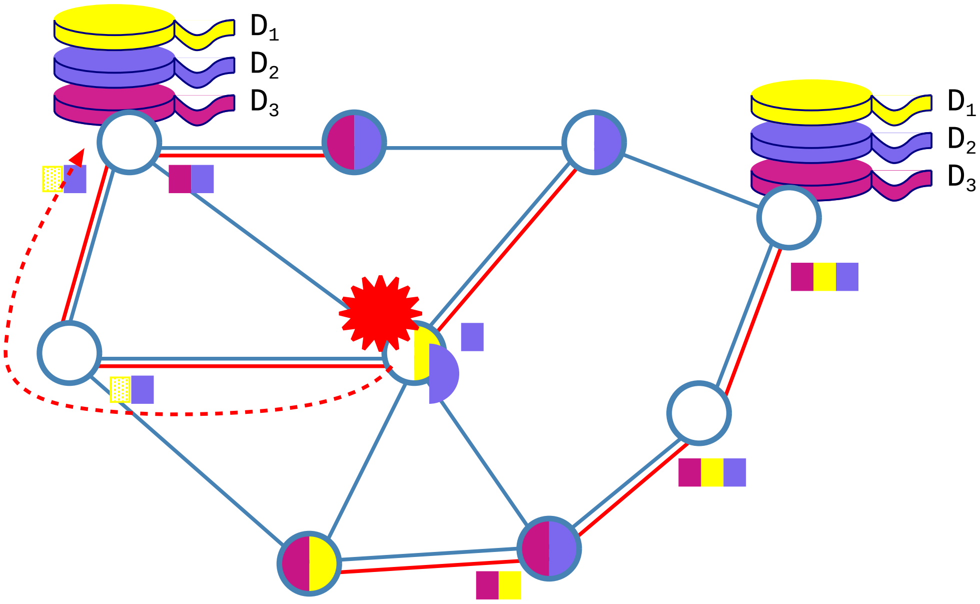


Exemple : groupes (1/2)

- Après trois étapes

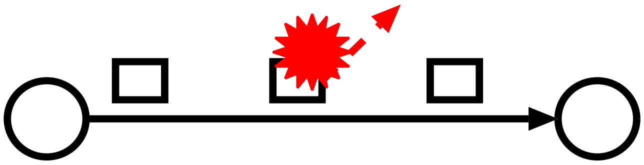


- Adaptativité aux groupes



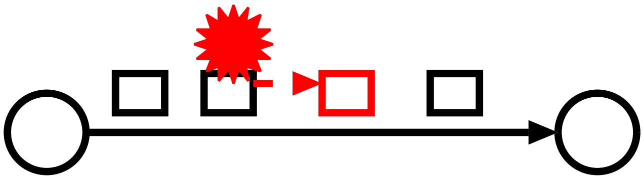
Communications non fiables

- Pertes



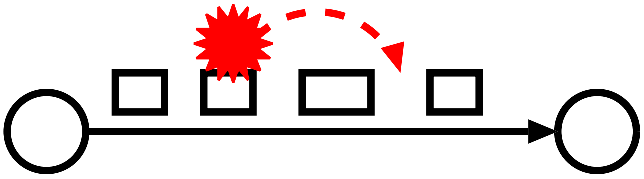
$\rightsquigarrow$ Retransmission sur temporisation

- Duplications



$\rightsquigarrow$ Idempotence de l'opérateur

- Déséquencement



$\rightsquigarrow$ Indépendance des messages

Conclusion

- Multidistribution multi-source adaptative
  - plusieurs sources, même information sur les sources
    choix de la meilleure source et du meilleur chemin
    utilisation de métriques variées
  - plusieurs unités de données $D_1 \ldots D_k$
    à chaque D_i correspond un groupe de destinataires G_i
    groupes dynamiques
    minimisation des diffusions dans la forêt
  - adaptation à l'infrastructure, au trafic...
    tolérance des défaillances transitoires
    communications non fiables

- Solution
  - composition d'un algorithme générique autostabilisant
  - paramétré par des r-opérateurs
  - utilisation de la preuve générique

$\rightsquigarrow$ algorithme simple pour la multidistribution multisources adaptative

$\rightsquigarrow$ mécanisme générique simple de construction de protocoles autostabilisants (tâches statiques)