Agréger Rapidement des Données est Difficile

Quentin Bramas, Sébastien Tixeuil

Agréger Rapidement des Données est Difficile †

Quentin Bramas[1,2] et Sébastien Tixeuil[1,2,3]

[1]*Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, F-75005, Paris, France*
[2]*CNRS, UMR 7606, LIP6, F-75005, Paris, France* [3]*Institut Universitaire de France*

La première utilisation des réseaux de capteurs sans fil est l'agrégation des données collectées par chaque capteur, de manière économe en énergie. Dans ce papier, nous présentons le problème de l'agrégation des données de manière formelle dans les réseaux de capteurs dynamiques (dont la topologie évolue avec le temps). Dans les réseaux statiques de degré au plus quatre, il a déjà été prouvé que résoudre ce problème en minimisant le délai est NP-Complet. Il en est donc de même dans les réseaux dynamiques.

Notre première contribution est de caractériser exactement ce qui fait de l'agrégation de données un problème NP-Complet. En effet, dans les réseaux *statiques*, où le problème est trivial pour les réseaux de degré au plus deux, nous montrons que le problème reste NP-Complet dans les réseaux de degré au plus trois. De même, le problème est trivial dans les réseaux *dynamiques* de degré au plus un et nous montrons qu'il est NP-Complet dans les réseaux dynamiques de degré au plus deux. Nous montrons donc que l'ajout de la dynamique rend le problème de l'agrégation de données intrinsèquement plus difficile. Notre deuxième contribution est de présenter des bornes inférieures et supérieures pour la durée d'agrégation des données dans un réseau dynamique. Là encore, l'ajout de la dynamique augmente la durée d'agrégation des données dans le pire des cas.

Keywords: Complexité, réseaux dynamiques, agrégation de données

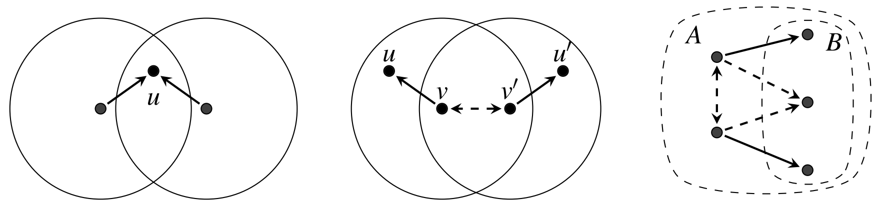
1 Introduction

Les réseaux de capteurs sans fil ou *Wireless Sensor Netowks* (WSN) sont constitués de nœuds capteurs, capables de récolter des données, de les analyser et de les transmettre. Ces réseaux ont des applications variées, en fonction de la zone où ils sont déployés : militaire ou de sauvetage dans des zones pouvant être inaccessibles aux humains ; sanitaire avec des capteurs déployés sur et dans le corps humain ; surveillance avec des capteurs sur les voitures d'une ville ou les arbres d'une forêt, etc. Les nœuds sont autonomes en énergie et il est primordial d'assurer leur longévité sans retarder la récolte des données.

On considère deux techniques principales pour réduire la consommation énergétique des capteurs. D'un côté, on évite les collisions produites par les transmissions simultanées de capteurs proches les uns des autres (voir figure 1). D'un autre côté, lorsqu'un capteur reçoit des données d'un capteur voisin, il les agrège à ses propres données avant de les retransmettre. Ainsi, une seule transmission sera suffisante pour transmettre les données agrégées. De cette manière n transmissions sont suffisantes pour récolter les données de n capteurs vers un noeud coordinateur (contre $\Omega(n^2)$ sans agrégation, dans le pire des cas). Le problème de l'agrégation des données consiste à trouver un ordonnancement des communications qui respecte ces deux règles et qui permet d'agréger les données de tous les capteurs du réseau jusqu'à un nœud coordinateur en un minimum de temps.

Modèle Un réseau de capteurs est un réseau *ad hoc*. À un instant donné, son graphe de communication (où les sommets représentent les capteurs et les arêtes représentent les liens de communication entre les capteurs) est un graphe d'intersections de disques de même rayon (représentant la portée de communication). On dit qu'un réseau de capteurs est dynamique si les capteurs sont mobiles. Les liens de communication d'un réseau dynamique peuvent disparaitre ou apparaitre au cours du temps, sans que cela soit considéré comme une erreur. Nous choisissons de modéliser un réseau de capteurs dynamique par un graphe *évolutif*,

†Ce travail a été effectué dans le cadre du Labex SMART soutenu par les financements de l'état Français gérés par l'ANR sous le programme Investissements d'Avenir sous la référence ANR-11-IDEX-0004-02. Aussi, il est soutenu en partie par le LINCS.



(a) u ne reçoit aucune donnée

(b) u reçoit les données de v et u' les données de v'

(c) Les données sont aggrégées de A vers B

FIGURE 1: Contraintes de Communication

c'est-à-dire une suite de graphes statiques. Chaque graphe de la suite représente le graphe de communication à un instant donné. De manière formelle, un graphe évolutif G est un couple $(V, \{E_i\}_{i\in\mathbb{N}})$ où E_i est l'ensemble des liens de communication au temps i. Dans le cas d'un réseau de capteurs sans fil, chaque graphe (V, E_i) est un graphe d'intersections de disques. Le degré maximal d'un graphe évolutif est le degré maximal parmi tous les graphe (V, E_i). Dans un graphe évolutif, une arête est un couple (a,t) où t est un temps et a une arête dans E_t. Ainsi, un chemin évoluant au cours du temps $\mathcal{T}$, appelé *trajet*, est une suite d'arêtes de la forme $((a_1,t_1),(a_2,t_2),\ldots,(a_l,t_l))$ qui commence au temps $depart(\mathcal{T}) = t_1$ et se termine au temps $arrivee(\mathcal{T}) = t_l + 1$.

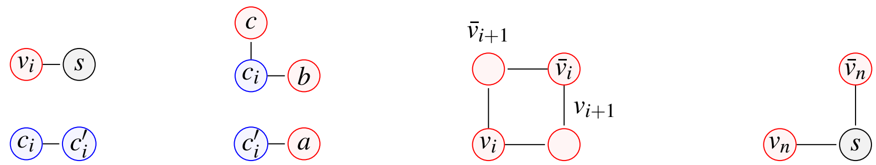
On dit que les données sont agrégées au temps t d'un ensemble de nœuds A vers un ensemble de nœuds $B \subset A$ si les données des nœuds de $A\backslash B$, transmises simultanément à l'instant t, sont correctement reçues par des nœuds de B (*i.e.*, sans interférence, voir figure 1c). C'est-à-dire que chaque nœud u de $A\backslash B$ possède au moins un voisin v dans B qui reçoit ses données (u est le seul voisin de v qui transmet au temps t). Une agrégation de données de V vers s de durée d et de départ t_0 est une suite décroissante d'ensembles $V = R_0 \supseteq R_1 \supseteq \ldots \supseteq R_d = \{s\}$ telle que les données sont agrégées au temps $t_0 + i$ de R_i vers R_{i+1}, pour tout $i = 0, 1, \ldots, d-1$.

Une instance du problème de l'*Agrégation de Données en Temps Minimum* (ADTM) est un triplet (G,s,t) où $G = (V, \{E_i\}_{i\in\mathbb{N}})$ est un graphe évolutif, $s \in V$ un nœud appelé *puits*, et t un temps. Une solution de l'instance $ADTM(G,s,t)$ est une agrégation de données de V vers s de départ t et de durée minimum. La durée minimum de l'agrégation est notée $ADTM_{Opt}(G,s,t)$. Dans le cas des graphes statiques, la donnée de t peut être omise.

Travaux connexes Dans le cas statique le problème de l'agrégation de données en temps minimum (ADTM) a été largement étudié. Une version similaire fut présentée par Anamalai *et al.* [AGS03] avec la possibilité pour les nœuds de choisir parmi un nombre fini de canaux différents pour éviter les collisions. Puis, Chen *et al.* [CHZ05] ont donné un modèle formel pour l'étude du problème ADTM, et ont montré que le problème est NP-Complet même si le graphe est de degré au plus quatre. Dans le même article, ils ont donné la borne inférieure et un majorant pour la durée minimum d'agrégation des données. Ils ont aussi proposé le premier algorithme d'approximation. Pour finir de nombreux articles ont présenté des algorithmes d'approximations centralisés et distribués de plus en plus performants, utilisant la nature géométrique des graphes d'intersections de disques, améliorant par la même occasion la majoration de la durée suffisante pour agréger les données du réseau. Cependant, aucune de ces études ne se généralise au cas des graphes évolutifs. À notre connaissance, aucune étude sur le problème de l'agrégation de données prenant en compte les collisions n'a été réalisée dans les graphes évolutifs.

Contribution Notre première contribution est de caractériser exactement ce qui fait de l'agrégation de données un problème NP-Complet. Dans les graphes *statiques*, le problème est trivial pour les graphes de degré au plus deux, et nous montrons que le problème reste NP-Complet dans les graphes de degré au plus trois. De manière similaire, le problème est trivial dans les graphes *dynamiques* de degré au plus un et nous montrons qu'il est NP-Complet dans les graphes de degré au plus deux.

Notre deuxième contribution est de donner les bornes inférieures et supérieures pour la durée de l'agrégation des données dans un graphe dynamique.



(a) configuration durant T_1 (b) configuration durant T_2 (c) configuration durant T_3 (d) dernière configuration de T_3

FIGURE 2: Configuration des noeuds (les clauses sont bleus et les littéraux sont rouges).

2 Complexité de l'Agrégation de Données en Temps Minimum

Dans les Réseaux de Capteurs Statiques Nous améliorons le résultat de Chen *et al.* [CHZ05] pour le rendre plus précis et optimal. Notre résultat s'applique aux graphes d'intersections de disques de rayon $1/2$ à coordonnées entières et de degré au plus trois (le résultat de Chen *et al.* est vrai dans les grilles partielles, qui sont un sur-ensemble strict des graphes d'intersections de disques que nous considérons, de degré au plus quatre). Par ailleurs, le problème est facile à résoudre dans le cas des graphes statiques de degré au plus deux. En effet, en posant ε l'excentricité du nœud puits et n le nombre de nœuds total, si n est impair et $\varepsilon = (n-1)/2$ (le graphe est soit un cycle, soit une ligne de centre s) alors $ADTM_{Opt}(G,s) = \varepsilon + 1$, sinon $ADTM_{Opt}(G,s) = \varepsilon$.

Ainsi, le résultat suivant est optimal en fonction du degré maximum du graphe.

Théorème 2.1 *Le problème ADTM est NP-Complet dans les réseaux de capteurs statiques de degré au plus trois.*

Dans les Réseaux de Capteurs Dynamiques Nous utilisons une approche différente pour démontrer la NP-Complétude dans les graphes évolutifs (qui modélisent les réseaux de capteurs dynamiques) de degré au plus deux. Encore une fois, le cas des graphes de degré au plus un est trivial étant donnée l'absence de collision.

Théorème 2.2 *Le problème ADTM est NP-Complet dans les réseaux de capteurs dynamiques de degré au plus deux.*

Preuve: La preuve se fait par réduction depuis le problème $3-$SAT. À partir d'une instance φ du problème $3-$SAT, avec n variables $v_1, \ldots, v_n$ et m clauses $c_1, \ldots, c_m$, on construit le graphe évolutif $G(V, \{E_i\}_{i\in\mathbb{N}})$ de la manière suivante. L'ensemble des nœuds est composé d'un nœud puits s, des littéraux de φ et des clauses de φ en deux exemplaires :

$$V = \{s\} \bigcup_{1\leq i\leq n} \{v_i\} \cup \{\bar{v}_i\} \bigcup_{1\leq i\leq m} \{c_i\} \cup \{c'_i\}$$

Soit $t_f = 3n+m$. On décompose l'intervalle de temps $[1, t_f]$ en trois périodes T_1, T_2 et T_3 (voir figure 2). Durant $T_1 = [1, 2n]$, on a pour tout $i \in [1, n]$:

$$E_{2i-1} = \left\{(v_i, s), \left(c_i, c'_i\right)\right\}, \quad E_{2i} = \{(\bar{v}_i, s)\}$$

Durant $T_2 = [2n+1, 2n+m]$, on a pour tout $j \in [1, m]$:

$$E_{2n+j} = \left\{\left(c'_j, a\right), (c_j, b), (c_j, c)\right\} \quad \text{avec } a, b \text{ et } c \text{ les littéraux de la clause } c_j$$

Durant $T_3 = [2n+m+1, 2n+m+n]$, on a pour tout $i \in [1, n-1]$:

$$E_{2n+m+i} = \{(v_i, v_{i+1}), (\bar{v}_i, v_{i+1}), (v_i, \bar{v}_{i+1}), (\bar{v}_i, \bar{v}_{i+1})\} \qquad \text{et} \qquad E_{2n+m+n} = \{(v_i, s), (\bar{v}_i, s)\}$$

Durant T_3, une variable ou sa négation (de manière exclusive) peut transmettre ses données au littéral suivant jusqu'au nœud puits, de sorte que l'ensemble des littéraux qui transmettent durant cette période, en leur assignant la valeur *vrai*, est vu comme une interprétation de la formule φ. On peut montrer que cette interprétation *satisfait* la formule φ si et seulement si la durée de l'agrégation des données de G vers s de départ 1 est t_f.

$\square$

3 Borne Inférieure et Borne Supérieure

Parmi tous les trajets qui commencent après l'instant t, un trajet $\mathcal{T}$ est dit *principal* s'il minimise la durée $arrivee(\mathcal{T}) - t$. Dans le cas statique, un trajet principal coïncide avec la notion de plus court chemin.

L'excentricité du nœud puits s peut être définie comme la hauteur commune des arbres enracinés en s de plus courts chemins. Elle est une borne inférieure dans le cas statique [CHZ05] car les arbres de plus courts chemins définissent autant de manières possibles d'agréger les données, lorsque les collisions ne sont pas prises en compte. C'est pourquoi, nous définissons un *arbre des trajets principaux* (ATP) enraciné en s de départ t comme étant un arbre enraciné en s ou chaque branche correspond à un trajet principal de la feuille vers s commençant après t. À la manière d'un arbre de plus courts chemins dans le cas statique, un ATP peut servir de base pour définir un ordonnancement pour l'agrégation des données dans un graphe évolutif. Ainsi, la durée commune des ATP, enracinés en s et de départ t, notée $ATPD(G,s,t)$, est une borne inférieure de $ADTM_{Opt}(G,s,t)$.

Dans le cas statique, un arbre de plus courts chemins reste toujours valide. Cela permet, pour résoudre une collision, de retarder d'une unité de temps une transmission. À chaque nœud d'un arbre de plus courts chemins, il y a au maximum $\Delta - 1$ collisions, où Δ est le degré maximum du graphe. D'où, $(\Delta - 1)\varepsilon$ est un majorant de $ADTM_{Opt}(G,s)$ [CHZ05]. Dans le cas dynamique, un ATP n'est valable que pour un temps donné, et rien ne certifie qu'une transmission peut être retardée d'une unité de temps car le lien de communication peut disparaitre. En fait, l'existence d'un ATP enraciné en s ne certifie pas l'existence d'une solution au problème ADTM.

Pour trouver une condition suffisante d'existence d'une solution, on remarque qu'un ATP permet au moins à un nœud donné de transmette sa donnée jusqu'au nœud puits. Une condition suffisante pour l'existence d'une solution au problème ADTM est alors la donnée de $(n-1)$ ATP consécutifs *i.e.*, une suite d'ATP dont la fin d'un ATP précède le début de l'ATP suivant. Cette observation permet aussi d'assurer que la durée de $(n-1)$ ATP consécutifs de départ t, notée $ATPD^{n-1}(G,s,t)$ est un majorant de $ADTM_{Opt}(G,s,t)$. Ce majorant est atteint par le graphe $G(V, \{V \times V\}_{i \in \mathbb{N}})$ de degré $n-1$ (et elle n'est pas atteinte par un graphe de degré maximum $n-2$ ou inférieur).

Théorème 3.1 *Soit G un graphe évolutif, s un nœud de G et t un temps, on a :*

$$ATPD^{n-1}(G,s,t) \geq ADTM_{Opt}(G,s,t) \geq ATPD(G,s,t)$$

4 Conclusion

Nous avons étudié le problème de l'agrégation de données dans les graphes évolutifs, qui modélisent les réseaux de capteurs dont la topologie évolue avec le temps. Nous avons donné les conditions optimales sur le degré maximum du graphe pour que le problème soit NP-Complet. De plus nous avons donné les premières bornes pour la durée d'agrégation des données. La borne inférieure est vraie quel que soit le degré maximum du graphe. Cependant, la borne supérieure n'est pas le plus petit des majorants pour les graphes de degré maximum $n-2$ ou inférieur. Des travaux futurs consisteraient à trouver une borne supérieure qui resterait optimale quel que soit le degré maximum du graphe.

Références

[AGS03] Valliappan Annamalai, Sandeep KS Gupta, and Loren Schwiebert. On tree-based convergecasting in wireless sensor networks. In *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, volume 3, pages 1942–1947. IEEE, 2003.

[CHZ05] Xujin Chen, Xiaodong Hu, and Jianming Zhu. Minimum data aggregation time problem in wireless sensor networks. In *Mobile Ad-hoc and Sensor Networks*, pages 133–142. Springer, 2005.