Sélection des brokers dans un réseau de capteurs en mode publication / souscription

Claude Chaudet[1], Nicola Costagliola[1], Isabelle Demeure[1], Salma Ktari[1] and Samuel Tardieu[1]

[1] *Institut Télécom / Télécom ParisTech CNRS LTCI UMR 5141 ; Paris, France ; {first.last}@telecom-paristech.fr*

Dans cet article, nous nous intéressons au déploiement, dans un réseau de capteurs sans-fil, d'une couche de communication de type publication et souscription s'appuyant sur un ensemble de nœuds intermédiaires, les *brokers* dont le nombre et la localisation a une influence sur la performance du système en termes de mémoire, d'énergie ou de délai d'acheminement. Nous comparons, par simulation numérique, la performance de plusieurs stratégies de placement et de la mise en pratique de ces critères. Ce travail a été réalisé dans le cadre du projet ANR Diaforus (VERSO 2009).

1 Introduction

Les réseaux de capteurs sont supposés à la fois être auto-organisants, efficaces en énergie et passer à l'échelle. Les nœuds qui les composent sont de faibles capacités et la conception d'un intergiciel efficace pour ces environnements est un réel défi. Dans ce contexte, l'asynchronisme du mode de transmission publication et souscription (*publish/subscribe*) est attrayant. Les données émises par les capteurs sont étiquetées avant d'être envoyées non pas à leur destination, mais à un intermédiaire, un *broker*, à qui les récepteurs potentiels ont exprimé leur intérêt sous forme d'expression sur les étiquettes. La correspondance et l'envoi des données aux souscripteurs incombe entièrement au(x) *broker(s)* et les émetteurs et les récepteurs n'ont besoin que d'une connaissance très limitée du réseau. En outre, cette correspondance engendre un filtrage des transmissions, améliorant l'efficacité énergétique du réseau s'il est réalisé proche des sources.

Il est possible d'avoir un ensemble quelconque de *brokers*, dont le nombre et l'emplacement vont influencer la performance du réseau. En nombre trop faible, les *brokers* seront des goulets d'étranglement et en nombre trop important ils produiront un trafic de contrôle important, par exemple pour le partage des tables de correspondance. Dans cet article, nous nous intéressons à la bonne façon de sélectionner cet ensemble de *brokers*. Au moyen d'un modèle de files d'attente, nous comparons plusieurs critères de sélection basés sur des métriques de graphes en termes d'efficacité énergétique de l'ensemble du réseau et en termes de charge imposée aux différents nœuds. Ces critères peuvent alors être utilisés dans un algorithme de *clustering* adapté que nous évoquons tout en discutant de la complexité de leur estimation ou approximation.

2 Aperçu de l'état de l'art

Si plusieurs travaux ont souligné l'importance de l'organisation de l'ensemble des *brokers* pour un système *publish/subscribe* dans l'Internet, peu de contributions ont examiné cette problématique dans le contexte particulier d'un réseau de capteurs. Dans [9], Tran et Truong décrivent un nombre considérable d'intergiciels à base de *publish/subscribe* conçus pour ces réseaux. Peu de travaux examinent l'ensemble des *brokers*. Quelques contributions présupposent qu'un seul *broker* central effectue le travail alors que d'autres supposent que les tâches des *brokers* sont réalisées par des nœuds élus sur d'autres critères. [7] par exemple décompose le réseau en zones au moyen d'un algorithme de *clustering k-means* et sélectionne le barycentre de chaque zone pour assumer ce rôle. [6] localise les *brokers* sur les multipoints relais du protocole de routage OLSR. Ces stratégies peuvent surcharger ou rendre critique certains nœuds et les critères qui influencent la sélection d'un rôle ou de l'autre sont *a priori* différents.

3 Modèle

Cette section présente et résume le modèle publié dans [1]. Nous nous plaçons au dessus de la couche réseau et supposons que la mobilité est suffisamment lente pour que le protocole de routage puisse l'absorber de façon transparente. Nous comparons un routage plus-courts-chemins avec un routage de type RPL créant et maintenant une structure de graphe acyclique dirigé enracinée au collecteur de données. Nous présupposons que la couche liaison a un fonctionnement similaire à IEEE 802.15.4 pour lequel l'émission d'une trame est assortie d'un surcoût fixe représentant un temps T_h et d'un surcoût variable. Elle peut engendrer la transmission d'un acquittement occupant le médium durant T_{ack}. Nous négligeons la mise en veille des capteurs, mais celle-ci peut être modélisée par un facteur multiplicatif sur certains termes.

Nous examinons tout d'abord l'**efficacité énergétique du réseau**, la consommation étant essentiellement due à la transmission radio. L'émission et la réception d'une trame provoquent des consommations comparables car il est nécessaire, pour recevoir une trame, d'extraire le signal du bruit. Considérons une version optimiste d'IEEE 802.15.4 dans laquelle les trames sont acquittées au niveau liaison par le récepteur et dans laquelle les voisins de l'émetteur d'une trame qui n'en sont pas le récepteur ne fassent que décoder l'en-tête de la trame et se mettent en veille pour la charge utile. Émettre une trame de longueur L bits à un débit D bit/s a alors un coût énergétique $E_e(L) = P_{tx}.(T_h + L/D) + P_{rx}.T_{ack}$, où P_{tx} représente la puissance (W) consommée à l'émission et P_{rx} à la réception. Le récepteur consomme une énergie $E_r(L) = P_{rx}.(T_h + L/D) + P_{tx}.T_{ack}$ et les voisins de l'émetteur une énergie $E_o = T_h.P_{rx}$.

Chaque publication (c-à-d paquet) est tout d'abord acheminée du publiant à un *broker*, *a priori* le plus proche de la source. Elle est ensuite transmise entre *brokers* et nous supposons, de façon optimiste, que la synchronisation des *brokers* est parfaite et que seuls les *brokers* auxquels sont associés un souscripteur intéressé la reçoivent avant de la transmettre auxdits souscripteurs. Un *broker* en charge de plusieurs souscripteurs ne recevra le paquet qu'une seule fois. Le coût de synchronisation reste raisonnable si le réseau ne comporte que peu de souscripteurs, ce qui est probable dans un réseau de capteurs. Nous comptabilisons alors chaque émission, réception ou écoute de trame afin de calculer l'énergie dépensée par chaque nœud.

Pour évaluer la **charge de chaque nœud**, nous faisons l'hypothèse que chaque publiant émet des trames selon un processus de Poisson d'intensité λ_p. En appliquant le schéma d'acheminement ci-dessus, il est possible d'évaluer l'intensité de trafic que chaque nœud du réseau aura à émettre. Si l'on considère le fonctionnement de la couche MAC, le temps de service à chaque saut sera constitué d'une part fixe et d'une part variable (*backoff*) et peut être modélisé par une variable aléatoire de moyenne $1/\mu$s et de variance σ. Un nœud peut être modélisé comme une file d'attente M/G/1 et le nombre de paquets en attente dans chaque nœud s'exprime en utilisant la formule de Pollaczek-Khinchin : $Q(i) = (1 + (\sigma.\mu)^2).\rho_i^2/(2.(1-\rho_i))$, où ρ est sa charge, somme des intensités de trafic qu'il doit acheminer divisée par le temps moyen d'émission d'une trame. En identifiant le nœud le plus chargé, il est possible de calculer l'intensité de trafic individuelle (λ_p) maximale ne conduisant pas à une instabilité. Ce chiffre représente, dans la suite, la **charge maximale tolérée par le réseau**. Enfin, à partir de l'expression de la taille de chaque file, il est possible d'évaluer, pour chaque publication, le **temps de remise maximal d'une publication** et par conséquent de limiter l'intensité de trafic afin de s'assurer que ce temps reste en deçà d'une limite donnée.

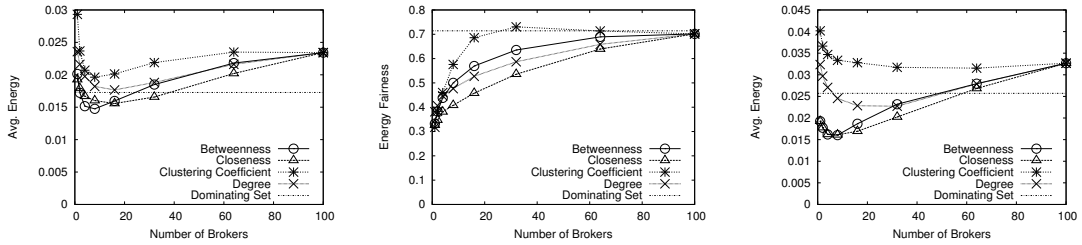
4 Comparaisons de plusieurs stratégies de sélection des *brokers*

Tout nœud du réseau, publiant souscripteur ou routeur, peut assumer le rôle de *broker*. La première stratégie de sélection des *brokers* que nous envisageons consiste à choisir comme *brokers* les nœuds constituant un ensemble dominant du graphe de connectivité $G = (V, E)$. Plusieurs algorithmes distribués ont été proposés à cet effet. Notons Δ le degré maximum de G ; [4] montre qu'un ensemble dominant approchant un ensemble minimal à un facteur $O(\log \Delta)$ peut être construit de façon distribuée en $O(\log |V| \log \Delta)$ étapes.

Nous comparons cette approche à une sélection des *brokers* au moyen de plusieurs métriques telles que la centralité d'intermédiarité (*betweenness centrality*), la centralité d'éloignement (*closeness centrality*), le degré et le coefficient de *clustering* [3]. Ces métriques sont souples car elles sont pondérables en fonction des rôles des nœuds et disposent d'algorithmes distribués d'approximation. La centralité d'intermédiarité peut être approchée par sa version égocentrique, calculée sur un voisinage à 2 sauts [8, 2]. La centralité d'éloignement peut être estimée à partir des tables de routage ou de la profondeur dans l'arbre de routage.

4.1 Quelques résultats

Dans ce travail nous cherchons uniquement à comparer les différentes métriques présentées ci-dessus. C'est pourquoi, à l'aide d'évaluations numériques réalisées par un outil développé à cet effet, nous évaluons la performance obtenues par ces différentes stratégies sur un ensemble de réseaux aléatoires dont le graphe de connectivité est un graphe géométrique aléatoire. Après avoir calculé les routes dans le réseau telles que le protocole de routage les obtiendrait, nous calculons pour chaque nœuds la valeur des différents critères et sélectionnons les n nœuds obtenant le meilleur score pour jouer le rôle de *brokers*, n étant l'un des paramètres que nous cherchons à caractériser. Nous évaluons ensuite la performance obtenue par ces différentes stratégies au moyen des modèles décrits précédemment en faisant varier n de 1 à $|V|$ de façon exponentielle ($n \in \{1,2,4,8,\ldots,|V|\}$). Les résultats présentés ici sont les moyennes obtenues sur une centaine de scénarios aléatoires, les publiants et souscripteurs étant répartis aléatoirement dans le réseau.



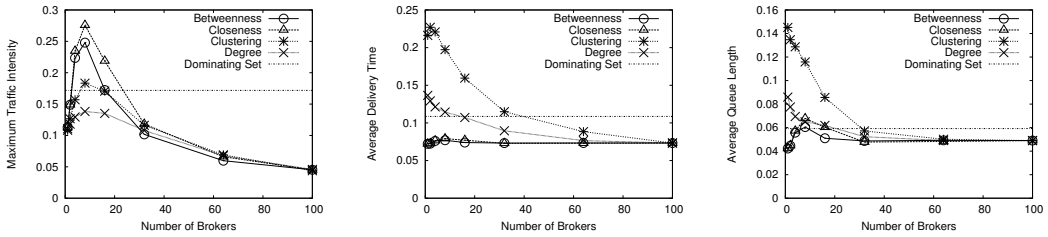
(a) Energie moyenne dépensée (W.h) - Plus courts chemins (b) Équité (critère de Jain) - Plus courts chemins (c) Energie moyenne dépensée (W.h) - routage arborescent (RPL)

FIGURE 1: Profil de consommation pour 90 publiants et 20 souscripteurs (plus-court-chemins)

Les figures 1(a) et 1(b) représentent l'énergie moyenne dépensée par chaque nœud et sa répartition (son équité) dans le cas d'un réseau comportant beaucoup de publiants et peu de souscripteurs. Pour cette première évaluation, les routes sont les plus-court-chemins et les valeurs de consommations correspondent à un circuit Chipcon CC2420. Aucun filtrage n'est effectué, mais une publication destinée à deux souscripteurs attachés au même *broker* n'est transmise qu'une fois dans le réseau de *brokers*. Nous pouvons remarquer que l'énergie est minimale, dans ce cas, pour un nombre assez faible de *brokers* et augmenter leur nombre est contre-productif alors même que nous ignorons, dans cette évaluation, les coûts liés au trafic de contrôle pour la maintenance et la synchronisation de l'ensemble des *brokers*. L'utilisation d'un ensemble dominant offre de bonnes performances, mais les centralités d'intermédiarité et d'éloignement arrivent à dépasser cette performance dans certains cas. Le degré et le coefficient de *clustering* tendent à diriger le trafic vers les zones denses et n'offrent pas de bonnes performances. Les résultats d'équité montrent que la consommation énergétique s'uniformise alors que le nombre de *brokers* augmente, ce qui tend à indiquer que les gains réalisés sur la moyenne ne concernent qu'un sous-ensemble des nœuds.

La Figure 1(c) est le pendant de la figure 1(a) lorsque le routage impose une structure d'arbre au réseau. On peut noter que si le profil reste similaire, l'utilisation d'un ensemble dominant présente une moins bonne performance alors que la performance obtenue en utilisant le degré augmente. Globalement, l'énergie dépensée augmente d'environ 50 % par rapport aux plus-court-chemins avec l'augmentation de la longueur des routes. Toutefois, les mesures de longueur des routes, non incluses ici, montrent qu'il n'y a pas de relation directe entre la longueur des routes et l'énergie consommée. L'équité (non représentée ici) est moins bonne avec RPL du fait du point de concentration que constitue la racine de l'arbre de routage.

Les Figures 2(a), 2(b) et 2(c) représentent respectivement l'évolution, dans le cas d'un routage arborescent, de la charge maximale tolérée par le réseau afin de maintenir la charge individuelle de chaque nœud en dessous de $0,9$, le temps moyen de remise d'une publication et la taille moyenne des files d'attente des nœuds. Ces deux dernières métriques sont calculées sur un réseau stable. L'ordre des métriques reste globalement le même, les centralités restant les plus performantes. Une fois la charge globale limitée, accroître le nombre de *brokers* au delà d'un certain seuil n'a que peu d'influence sur le délai ou la charge.



(a) Charge maximale tolérée par le réseau (pkt/sec)

(b) Temps moyen de remise des publications à l'ensemble des souscripteurs (s)

(c) Taille moyenne de file d'attente (paquets)

FIGURE 2: Métriques liées à la charge - réseau avec routage type RPL - 90 publiants et 20 souscripteurs

D'autres expérimentations montrent que le nombre de publiants dans le réseau définit le niveau de performance global mais ne change pas l'ordre entre les métriques. En revanche, lorsque le nombre de souscripteur est très faible (1 ou 2), la meilleure stratégie (en excluant les coûts de synchronisation) consiste à ce que chaque nœud joue le rôle de *broker*. Lorsque ce nombre augmente, la tendance observée sur les premiers résultats s'accentue, l'optimum étant plus marqué et situé autour de 16 *brokers* pour un réseau de 100 nœuds lorsque RPL est utilisé. Dans le cas où les plus courts chemins sont calculés entre tout couple de nœuds, l'approche basée sur un ensemble dominant se révèle très efficace. Augmenter la densité ou le diamètre du réseau ne modifie pas les conclusions.

5 Conclusion et extensions

Dans cet article, nous avons comparé la performance d'un réseau de capteurs fonctionnant en mode publication / souscription en comparant plusieurs critères de sélection des *brokers*. Trois critères ou approches se dégagent du lot, la première consistant à confier cette tâche à un ensemble dominant offre des performance raisonnables mais peut être surpassée par l'utilisation de métriques de centralité. Pour compléter ce travail, il faut intégrer ces critères dans un algorithme distribué de sélection, dérivant par exemple des algorithmes auto-stabilisants de *clustering* tels que [5] en adaptant notamment le critère permettant d'ôter un nœud de l'ensemble des *brokers* pour aller au delà des aspects topologiques. Cette adaptation permettra d'évaluer le facteur d'approximation introduit par le contexte distribué, notamment sur le nombre de *brokers* effectivement sélectionnés, ainsi que d'évaluer les surcoûts en communication.

Références

[1] C. Chaudet, I. Demeure, and S. Ktari. A model to evaluate brokers overlays for publish/subscribe in wireless sensor networks. In *WONS 2012*, Courmayeur, Italy, Jan. 2012.

[2] M. Everett and S. P. Borgatti. Ego network betweenness. *Social Networks*, 27, 2005.

[3] L. C. Freeman. Centrality in social networks : Conceptual clarification. *Social Networks*, 1(3), 1979.

[4] L. Jia, R. Rajaraman, and R. Suel. An efficient distributed algorithm for constructing small dominating sets. In *PODC*, 2001.

[5] C. Johnen and L. H. Nguyen. Robust self-stabilizing clustering algorithm. In *OPODIS 2006*, Bordeaux, France, Dec. 2006.

[6] A. R. Khakpour and I. Demeure. Chapar : A Cross-Layer Overlay Event System for MANETs. In *MobileWireless Middleware, Operating Systems, and Applications*, 2009.

[7] Y. Liu, B.-C. Seet, and A. Al-Anbuky. Virtual brokers for large-scale publish/subscribe in wireless sensor networks. In *IEEE/IFIP EUC '10*, Hong Kong, China, Dec. 2010.

[8] P. V. Marsden. Egocentric and sociocentric measures of network centrality. *Social Networks*, 24, 2002.

[9] D. A. Tran and L. H. Truong. *Advances in Next Generation Services and Service Architectures*, chapter Chapter 36 - Enabling Publish/Subscribe Services in Sensor Networks. River Publishers, 2011.