

Systèmes d'information pair-à-pair pour les réseaux de capteurs larges échelles

Yann Busnel (Projet Paris)
email : Yann.Busnel@irisa.fr

Thèse financée par le Ministère de la recherche

1 Introduction

L'évolution de l'informatique a été marquée par des étapes dans la miniaturisation. Dernièrement, sont apparus d'infimes systèmes micro-électromécaniques notamment les réseaux de capteurs sans fil (WSN). De nombreuses applications nouvelles pourraient émerger mais il existe encore des défis. Ces réseaux se répandent du fait de leur faible coût, de leur adaptabilité et de leur composition de dispositifs compacts et à faible consommation (cf. figure 1). Ces derniers, appelés communément *poussières intelligentes* ou *capteurs*¹, sont constitués d'une plate forme matérielle couplée avec un module d'acquisition permettant de capter et collecter des événements, d'analyser les traitements et de transmettre les informations recueillies dans divers environnements. Outre les capteurs eux-mêmes, les WSN possèdent le plus souvent des stations de contrôle appelées *nœuds-puits* ou *station de base*. Ceux-ci possèdent plus de ressources matérielles et permettent de collecter et stocker les informations issues des capteurs.

L'intérêt de la recherche et de l'industrie dans ce domaine s'intensifie par la potentielle fiabilité, précision, flexibilité, le faible coût ainsi que la facilité de déploiement de ces systèmes. Les caractéristiques des WSN permettent d'envisager de nombreuses applications d'observation réparties dans l'espace. Ces dernières peuvent se déployer dans de nombreux contextes : observation de l'environnement naturel (pollution, inondation, ...), suivi d'écosystèmes (surveillance de niches d'oiseaux, croissance des plantes, ...), militaire (télésurveillance de champs de bataille, détection d'ennemis, ...), biomédical et surveillance médicale (détection de can-

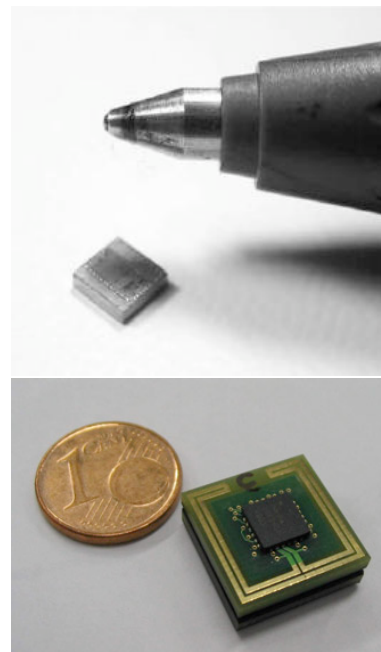


FIG. 1 – Exemples de capteurs

¹Nous choisissons cette dénomination dans le reste de cet article.

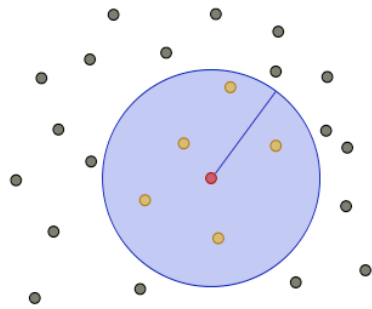


FIG. 2 – La vision d'un capteur dépend du rayon de réception de son module de communication.

cer, rétine artificielle, taux de glucose, diabète, ...), etc.

La spontanéité, l'adaptabilité du réseau et la dynamique de la topologie dans le déploiement des WSN soulèvent de nombreuses questions ouvertes. De plus, compte tenu du nombre potentiellement important de capteurs participants et de leurs ressources réduites, il apparaît indispensable de développer des solutions entièrement décentralisées, répartissant la charge entre les participants et permettant de passer à l'échelle. De nombreux procédés des systèmes distribués, et plus particulièrement ceux issus du paradigme pair-à-pair (*P2P*), nous apparaissent comme adaptables aux réseaux de capteurs sans fil.

Le paradigme de communication pair-à-pair en particulier possède une capacité intrinsèque de passage à l'échelle. Les systèmes reposant sur celui-ci (dit *systèmes P2P* par la suite) sont totalement décentralisés, chaque nœud jouant à la fois le rôle de client et de serveur. La charge étant répartie sur tous les nœuds du système, augmenter le nombre de participants n'induit pas de goulot d'étranglement ni de point critique de défaillance, couramment observés lors de l'utilisation du paradigme *client/serveur*. De plus, la vision uniquement locale des nœuds des systèmes *P2P* (*ie.* pas de connaissance globale du système, communication avec un sous ensemble restreint de participants – cf. figure 2) rendent l'adaptation de ceux-ci aux WSN avantageuse.

2 Réseaux de capteurs

Alors même qu'un grand nombre d'applications mettent en jeu des WSN, ceux-ci ont plusieurs restrictions que ces applications doivent contourner. Par exemple, ils ont une faible puissance de calcul, une réserve d'énergie limitée et une bande passante réduite aux connexions sans fil entre capteurs. Un des principaux objectifs dans la conception de WSN est la transmission de données fiable *via* une heuristique de préservation d'énergie et de prévention de perte de connectivité (*e.g.* aucun nœud isolé), ceci par l'utilisation de politique stricte de gestion d'énergie. En effet, la principale source de consommation d'énergie d'un capteur est l'utilisation du réseau sans fil *via* son module de radiocommunications. Les protocoles de routage² au sein des WSN sont influencés par plusieurs facteurs déterminants. Nous présentons tout d'abord trois défis communs à tous les WSN, puis un ensemble de modèles portant sur un type particulier d'application [1] :

Consommation d'énergie sans perte d'efficacité

Les capteurs utilisent leur réserve d'énergie à des fins de calcul et de transmission de données. La durée de vie d'un capteur dépend essentiellement de celle de sa batterie [2]. Dans un WSN, chaque nœud joue le rôle d'émetteur et de routeur³. Une défaillance énergétique d'un capteur peut changer significativement la topologie du réseau et imposer une réorganisation coûteuse de ce dernier.

Tolérance aux fautes Certains nœuds peuvent être défaillants ou inhibés à cause d'un manque d'énergie, d'un dommage physique ou d'une interférence. Ceci ne doit pas affecter la globalité de la tâche du réseau de capteurs. En cas de défaillance, de nouveaux liens et routes doivent être établis pour assurer la collecte des données. La redondance peut également être utilisée, tout en veillant à conserver une faible consommation d'énergie.

Couverture Dans les WSN, chaque nœud perçoit une vue locale de son environnement, limitée par sa portée et sa précision (cf. figure 2). La couverture d'un vaste espace déterminé est

²Ensemble de règles permettant de définir un chemin entre une source et sa destination

³Retransmission d'un message en provenance d'un émetteur distinct vers un destinataire donné

donc composée de l'union de nombreuses couvertures de petite taille.

Déploiement de noeuds C'est un facteur dépendant de l'application qui affecte grandement les protocoles de routage. Le déploiement peut être déterministe ou aléatoire. Dans la première stratégie, les capteurs sont placés manuellement et les données peuvent donc être acheminées via des chemins prédéterminés. En revanche, avec une approche aléatoire, les capteurs sont éparpillés (*e.g.* lâchés d'un avion). Dans le cas d'une répartition non uniforme, une stratégie de groupement (*clustering*) peut s'avérer nécessaire.

Modèle de couverture de données La capture d'information et la couverture des données requise dépendent essentiellement de l'application. Elles peuvent être classifiées selon différents modèles : en fonction du temps (surveillance périodique), des événements (réaction à l'occurrence d'un événement particulier), des requêtes (réponse à une demande d'une station de base) ou de manière hybride (combinaisons des précédentes approches) [3].

Hétérogénéité des noeuds / connections Dans de nombreuses études, tous les capteurs d'une application sont considérés homogènes (*i.e.* même capacité de calcul, de communication et d'énergie). Néanmoins, en fonction de l'application, certains capteurs peuvent avoir des rôles différents, générant une architecture hétérogène. Par exemple, dans un WSN hiérarchique, certains capteurs sont déclarés « chef » de leur groupe. Le routage vers les stations de base est alors traité par ces derniers.

Passage à l'échelle Le nombre de capteurs déployés peut être de l'ordre de la centaine ou du millier, et vraisemblablement plus dans un futur proche. Les algorithmes de routage doivent être capables de fonctionner efficacement avec une grande quantité de capteurs. De plus, ces algorithmes doivent traiter un grand nombre d'événements sans être saturés.

Dynamisme du réseau La plupart des architectures réseau reposent sur des capteurs statiques. Pourtant, la mobilité des stations de base et/ou des capteurs est parfois nécessaire dans de nombreuses applications (*cf.* section 4). La transmission de messages en provenance, ou vers, un nœud mobile est un autre défi. La cap-

ture peut être aussi bien statique que dynamique, dépendant de l'application (traque d'ennemis vs. surveillance de feux de forêt).

Connectivité La plupart des WSN possèdent une densité importante de capteurs, excluant ainsi l'isolement de nœuds. Pourtant, le déploiement, la mobilité et les défaillances font varier la topologie du réseau, dont la connectivité n'est pas toujours assurée.

Agrégation de données De nombreux capteurs peuvent générer de la redondance dans les mesures effectuées. L'agrégation de paquets similaires en provenance de différents nœuds permet de réduire le nombre de transmissions (*e.g.* suppression de duplicata, minimum, moyenne, maximum, ...) Un traitement du signal peut également être effectué pour l'agrégation, par fusion de données (*e.g.* conformation de faisceaux, ou *beamforming*).

Qualité de service Dans diverses applications, la donnée doit être transmise dans une certaine plage de temps, après quoi, elle devient caduque. Pourtant, dans la plupart des applications, la durée de vie du réseau est favorisée au détriment de la qualité d'émission des données. Les protocoles de routage assurant une qualité de service et prenant en compte la gestion de l'énergie, représentent un défi nouveau et stimulant.

3 L'approche pair-à-pair

3.1 Paradigme

De nombreuses approches permettant de construire un système issu du paradigme pair-à-pair ont récemment émergé, formant des systèmes distribués de grande taille ayant la capacité de passer à l'échelle. Contrairement aux approches *client-serveur*, les nœuds sont connectés directement entre eux par des liens logiques⁴ et jouent à la fois potentiellement les rôles de client et de serveur. Ces systèmes permettent notamment d'agréger les ressources et les services de tous les nœuds du système. Chaque nœud ne possède qu'une connaissance locale du système, permettant ainsi de développer des applications ayant des propriétés

⁴au-dessus du réseau physique

d'auto-organisation et de tolérance aux défaillances particulièrement adaptées à la grande taille des nouveaux systèmes distribués.

3.2 Applications

Les architectures pair-à-pair sont utilisées au sein d'un grand nombre d'applications, telles que :

Communication et collaboration Cette classe d'applications regroupe les systèmes fournissant une infrastructure de communication et de collaboration directe entre les différents acteurs du réseau, le plus souvent en temps-réel (*e.g.* messagerie instantanée telle que Jabber [4])

Calcul distribué L'objectif de ces systèmes est de tirer profit des ressources inutilisées des nœuds du réseau (*e.g.* mémoire, temps processeur, ...). Un processus nécessitant un calcul considérable est divisé en de nombreuses petites tâches, chacune répartie sur les différents nœuds du réseau. Une unité centrale de coordination des tâches est le plus souvent utilisée pour diviser, répartir et collecter les résultats (*e.g.* Seti@home, genome@home [5], ...)

Support de service pour Internet Un grand nombre d'applications issues des infrastructures pair-à-pair ont émergé, proposant divers services internet (*e.g.* diffusion multipoint [6], gestion des indirections, protection contre les dénis de service, ...)

Système de base de donnée D'importants travaux ont été menés sur la conception de base de données distribuées. Bernstein et al. ont proposé le modèle relationnel local (LRM) [7], dans lequel l'ensemble des données est stocké sur un réseau pair-à-pair en de multiples bases de données locales, possédant des dépendances sémantiques entre elles. PIER [8] est un moteur de recherche large échelle, construit au-dessus d'un réseau pair-à-pair. De nombreuses propositions existent telles que Piazza [9] (infrastructure de web sémantique) ou Edutella [10] (meta-données et système de requêtes pour applications P2P).

Distribution de données La plupart des systèmes P2P font partie de cette classe d'application, laquelle comprend les systèmes et infrastructures conçus pour le partage de médias numériques et autres données entre les utilisateurs.

La distribution de données P2P balaie un large spectre, partant d'un simple système de partage de fichier direct jusqu'à des systèmes complexes construisant une infrastructure de stockage distribuée permettant la publication, l'organisation, l'indexation, la recherche, la mise à jour et la récupération de données (*e.g.* Ka-ZaA [11], Oceanstore [13], Past [14], CAN [12], Mnemosyne [15], ...).

4 Objectifs et axes de recherche

La mise en œuvre des techniques pair-à-pair pour les réseaux de capteurs est un axe de recherche stimulant, prometteur et non trivial. En effet, les caractéristiques de connexions et de vision locale permettent de mettre en évidence les similarités de ces deux domaines. D'autre part, les faibles ressources des capteurs et leur voisinage imposé par leur portée de communication rendent difficile une application directe. Plusieurs facteurs entrent en jeu comme l'adaptation des méthodes P2P aux contraintes énergétiques, et ainsi, une maximisation de la durée de vie du réseau. Les systèmes P2P sont de manière inhérente auto-organisables, résistants aux défaillances et capables de passer à l'échelle.

De nombreux axes de recherche émergent à la frontière de ces deux domaines. Nous nous focalisons actuellement sur les objectifs suivants :

Diffusion La diffusion d'information dans les WSN se fait souvent par inondation (*flooding*) induisant une forte consommation d'énergie, surtout autour du point d'entrée de la donnée diffusée. L'utilisation de protocole épidémique P2P [16] est une solution envisageable pour diffuser de manière équitable l'information tout en conservant une rapidité de convergence.

Agrégation Plusieurs techniques d'agrégation de données sont déjà existantes dans les réseaux de capteurs, mais l'utilisation de techniques P2P dans l'agrégation en environnement à forte mobilité est un objectif encourageant.

Collecte En environnement dynamique et mobile, sans matériel de localisation précise des nœuds (*e.g.* GPS), l'envoi des données vers la station de base est coûteuse car le plus souvent effectuée par inondation. Une possibilité serait d'intégrer une information sur chaque nœud, mise

à jour de manière épidémique, permettant d'estimer la position de la station de base sans mécanisme coûteux (en temps, en énergie, ...)

Base de données physique Un intérêt émergent dans la communauté des réseaux de capteurs est la mise en place de bases de données physiques. Dans cette application, chaque capteur représente une entrée de la base de données, et les requêtes sur la base se traduisent par une collecte et/ou une agrégation de d'information provenant des nœuds participants. Dans ce cas, une base de données n'est plus la mutualisation de données physiques, mais ces dernières elles-même forment la base de donnée. Par exemple, dans la gestion des stocks d'un grand magasin, le décompte des produits serait automatique, et continuellement à jour.

5 Conclusion et perspectives

Domaine émergent, aux frontières de l'étude des systèmes informatiques, des réseaux numériques et du traitement du signal, l'attrait pour les réseaux de capteurs sans fil s'accroît considérablement. Nous avons présenté les propriétés souhaitées de ces réseaux, tout en mettant en évidence leurs limitations (physiques, technologiques, ...).

L'intégration du paradigme pair-à-pair aux WSN est un objectif attrayant et prometteur, dû en majeure partie aux similarités existant entre les deux approches.

Cette étude représente l'axe principal de mes travaux pour ma thèse de Doctorat. L'utilisation de systèmes P2P pour la diffusion épidémique et la collecte d'information en milieu mobile sont deux objectifs avancés en collaboration avec Marin Bertier (IRISA / INSA Rennes), Eric Fleury (CITI / INSA Lyon), Aline C. Viana (IRISA / INRIA), sous la direction d'Anne-Marie Kermarrec (IRISA / INRIA).

Références

- [1] J.N. Al-Karaki et A.E. Kamal, *Routing Techniques in Wireless Sensor Networks : a Survey* *Wireless Communications*, IEEE, Volume 11, Issue 6, On pages 6-28, ISSN : 1536-1284, December 2004.
- [2] W. Heinzelman, A. Chetarakasan et H. Balakrishnan, *Energy-Efficient Communication Protocol for*

Wireless Microsensor Networks In Proceedings of the 33rd Hawaii International Conference on System Science (HICSS'00), January 2000.

- [3] Y. Yao et J. Gehrke, *The cougar approach to in-network query processing in sensor network* In SIGMOD Record, September 2002.
- [4] The Jabber web site. <http://www.jabber.org/>, Accessed on-line 2003.
- [5] S.M. Larson, C. Snow et V.S. Pete, *Modern Methods in Computational Biology* In chapter Folding@Home et Genome@Home : Using distributed computing to tackle previously intractable problems in computational biology. Horizon Press, 2003.
- [6] M. Castro, P. Druschel, A.-M. Kermarrec et A. Rowstron. *Scribe : A large-scale et decentralized application-level multicast infrastructure*. In IEEE Journal on Selected Areas in Communications, 20(8), October 2002.
- [7] P. Bernstein, F. Giunchiglia, A. Kementsietsidis, J. Mylopoulos, L. Serafini et I. Zaihrayeu. *Data management for peer-to-peer computing : A vision*. In Proceedings of the Workshop on the Web et Databases, WebDB'02., 2002.
- [8] R. Huebsch, J.M. Hellerstein, N. Lanham et B. Thau Loo. *Querying the internet with pier*. In Proceedings of the 29th VLDB Conference, Berlin, Germany, 2003.
- [9] A.Y. Halevy, Z.G. Ives, P. Mork et I. Tatarinov. *Piazza : Data management infrastructure for semantic web applications*. In Proceedings of the twelfth international conference on World Wide Web, pages 556-567, Budapest, Hungary, 2003.
- [10] W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmer et T. Risch. *Edutella : A p2p networking infrastructure based on rdf*. In Proceedings of the twelfth international conference on World Wide Web, Budapest, Hungary, 2003.
- [11] The Kazaa web site. <http://www.kazaa.com>, Accessed on-line 2003.
- [12] S. Ratnasamy, P. Francis, M. Hetley, R. Karp et S. Shenker. *A scalable content-addressable network*. In Special Interest Group on Data Communications (ACM SIGCOMM'01), San Diego, CA, USA, Aug. 2001.
- [13] J. Kubiawicz, D. Bindel, Y. Chen, P. Eaton, D. Geels, S.R. Gummadi, H. Weatherspoon, W. Weimer, C. Wells et B. Zhao. *Oceanstore : An architecture for global-scale persistent storage*. In Proceedings of ACM ASPLOS. ACM, November 2000.
- [14] A. Rowstron et P. Druschel. *Pastry : Scalable, distributed object location et routing for large-scale peer-to-peer systems*. In Proceedings of IFIP/ACM Middleware, Heidelberg, Germany, November 2001.

- [15] S. Het et T. Roscoe. *Mnemosyne : Peer-to-peer steganographic storage*. In Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02), MIT Faculty Club, Cambridge, MA, USA, March 2002.
- [16] M. Jelasity, R. Guerraoui, A.-M. Kermarrec et M. van Steen. The peer sampling service : Experimental evaluation of unstructured gossip-based implementations. In *ACM/IFIP/USENIX 5th International Middleware Conference (Middleware'04)*, Toronto, Ontario, Canada, October 2004.