

# Agrégation robuste de données massives à la volée : application aux compteurs électriques communicants

Benoît Grossin\*, Yousra Chabchoub\*\*

\*EDF R&D, 1 avenue du Général de Gaulle 92140 Clamart  
benoit.grossin@edf.fr

\*\* ISEP, 28 Rue Notre Dame des Champs 75006 Paris  
BILab, Télécom ParisTech, 46 rue Barrault 75634 Paris Cedex 13  
yousra.chabchoub@isep.fr

**Résumé.** Dans les années à venir, plusieurs millions de compteurs électriques communicants seront déployés sur l'ensemble du territoire français. Afin d'assurer la fiabilité d'un réseau de cette envergure nous proposons une topologie de communication multi-chemins qui repose sur la duplication des données transmises. Toute exploitation des données collectées doit alors tenir compte de la présence d'éléments dupliqués. Dans cet article, nous proposons une nouvelle méthode permettant de calculer en ligne des consommations électriques agrégées (agrégation spatiale). L'idée est d'adapter l'algorithme probabiliste *Summation sketch* de Considine et al. au contexte des compteurs communicants. Cette approche a l'avantage d'être insensible à la duplication et permet de profiter de la structure massivement distribuée du réseau de communication des futurs compteurs électriques. L'expérimentation de cette méthode sur des données réelles montre qu'elle donne une bonne précision sur l'estimation des consommations agrégées. Cette approche est aussi complétée par une méthode basée sur la théorie des sondages : On obtient une meilleure réactivité de l'estimateur avec rapidement et donc sur des données significativement partielles une erreur inférieure à 2.5%.

## 1 Introduction

Les pannes et les pertes de paquets sont très fréquentes dans les réseaux sans fil à cause des problèmes d'interférence, de la collision entre les paquets et de la faible puissance du signal. Plusieurs techniques ont été développées et mises en oeuvre pour limiter la perte de données et garantir une certaine fiabilité dans les transmissions. On peut citer par exemple les codes correcteurs qui sont essentiellement basés sur la redondance (voir par exemple (Caire et al. (1998))), ou encore la duplication. Dans ce papier, on se limite à l'étude de la duplication. Il s'agit d'envoyer la donnée plusieurs fois afin d'augmenter ses chances d'arriver au moins une fois à la destination. Suite à cet envoi multiple, tous les cas sont envisageables : le meilleur des cas est que la destination reçoive exactement une fois le message. Le pire des cas est que le message soit perdu malgré la duplication. Enfin, le cas le plus fréquent est que le message