

Conception de fonctions d'agrégation de données par apprentissage fédéré pour une architecture IoT Edge/Cloud

Laboratoire d'accueil : Laboratoire de Traitement et Transport de l'Information (L2TI) / Équipe Réseaux

Co-directeurs de thèse :

- Gladys Diaz (MCF HDR) gladys.diaz@univ-paris13.fr
- Khaled Boussetta (MCF HDR) khaled.boussetta@univ-paris13.fr

Contexte et motivation : L'objectif de cette thèse est de contribuer aux travaux de conception d'une architecture IoT, destinée à collecter des données permettant d'évaluer l'exposition des individus à des facteurs nuisibles pour la santé comme la pollution ou le risque de contamination bactériologique à l'intérieur des lieux publics, tels que les hôpitaux, les écoles ou les résidences pour personnes âgées.

L'infrastructure est constituée de capteurs à bas coût hétérogènes (polluants, conditions météorologiques, de présence etc.) qui collectent des données de types et de formats variées. Ces données sont transmises par liaisons radio à des nœuds de bordure (*edge*). Chaque nœud de bordure héberge des serveurs qui vont traiter localement les données afin d'assurer une collecte participative sécurisée et anonyme. Ensuite, chaque nœud de bordure transmet vers le cloud une quantité agrégée de données.

Les capteurs utilisés ne sont pas forcément très précis et ont des ressources de traitement, de transmission et d'autonomie énergétique relativement limitées. Par ailleurs, l'évaluation de l'exposition individuelle en intérieur à des facteurs nuisibles pour la santé requière la collecte de données variées, multi-échelles dans le temps et dans l'espace et une certaine connaissance du contexte local (e.g. activités humaine et industrielle). Or, encore aujourd'hui, les capteurs sont le plus souvent programmés pour transmettre des données de façon régulière et agnostique du contexte environnant. Ce mode de fonctionnement n'est évidemment pas optimal.

Problématique scientifique et contributions attendues : Nous proposons d'étudier dans le cadre de cette thèse l'utilisation des techniques d'apprentissage fédéré dans les nœuds de bordure pour concevoir des fonctions d'agrégation temporelle et spatiale des données. Certaines fonctions seront utilisées par les capteurs pour réguler l'envoi des données vers les nœuds de bordure. D'autres fonctions d'agrégation seront implémentées dans les nœuds de bordures pour définir les données et les stratégies de transmission vers le cloud.

Nous proposons d'utiliser des techniques d'apprentissage fédéré pour permettre aux nœuds de bordure d'élaborer de façon autonome les fonctions d'agrégation adéquats. Ces fonctions seront déployées et dimensionnées de façon autonomes en utilisant les technologies récentes de virtualisation des réseaux.

Outre l'optimisation des ressources, les fonctions d'agrégation doivent préserver la qualité des données collectées tout en assurant la sécurité et l'anonymat des sources. Elles doivent être élaborées en tenant compte des ressources de l'ensemble de l'infrastructure, du type de données et du contexte environnant, qui est multi-échelle et dynamique.

Méthodologie : La première étape consistera à réaliser un état de l'art sur les architectures IoT destinées à mesurer les facteurs nuisibles à la santé en environnement intérieur, les techniques d'agrégation spatiale et temporelle [1] et les travaux concernant l'orchestration autonome des fonctionnalités réseaux virtualisées [2].

Nous nous intéresserons ensuite à l'étude de fonctions d'agrégation issues des techniques d'apprentissage fédéré. Ces travaux s'appuieront sur des données collectées par la plateforme IoT conçue au L2TI, ainsi que celles partagées par une équipe de chercheurs du laboratoire (IPICYT) au Mexique, avec qui nous collaborons sur les problématiques de mesure de la pollution intérieure dans les bâtiments publics.

Les performances seront évaluées en utilisant l'infrastructure de la plateforme IoT développée par le L2TI ainsi que les ressources de calcul offertes par le grid MAGI.

Références :

- [1] Jin Cui, Khaled Boussetta, Fabrice Valois, **Classification of data aggregation functions in wireless sensor networks**, Computer Networks, Volume 178, 2020, doi: 10.1016/j.comnet.2020.107342.
- [2] M. Chahbar, G. Diaz, A. Dandoush, C. Cérin and K. Ghoumid, **A Comprehensive Survey on the E2E 5G Network Slicing Model**, in *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 49-62, March 2021, doi: 10.1109/TNSM.2020.3044626.