

A Multi-Scale Algorithm for Joint Forecasting-Scheduling to Solve the Massive Access Problem of IoT

Fiche de lecture

Hervé Mbilo

promoteur

Marie-Ange Remiche

2023-01-01

1 Introduction

L'internet des objets à un impact considérable sur nos vies, et-ce à plusieurs niveaux. Nous les portons sur nous, sous forme de montres, de capteurs biométriques ou physiologiques, de lunettes ou casque à réalité virtuelle et de vêtements connectés. Ces technologies portables ont pour objectif de nous aider à améliorer notre santé, notre bien-être ainsi que notre sécurité. Dans nos maisons, la domotique se développe à une vitesse grand V. Aujourd'hui, plus besoin de se soucier de l'approvisionnement alimentaire, car le réfrigérateur sait quand il n'y a plus assez de lait et le cellier de sucre ou de café. Dès lors, un signal est envoyé à un robot majordome qui effectue la commande à la boutique adéquate. Outre, la gestion de stock alimentaire, ce dernier peut orchestrer différents travaux ménagers tels que le repassage, le nettoyage et la tonte de gazon, en gérant les actions des robots domestiques (aspirateurs, robot de repassages, etc). A l'échelle d'une ville, on assiste à l'évènement des voitures autonomes. Ces véhicules intelligents peuvent communiquer entre eux afin de garantir la fluidité et la sécurité du trafic routier.

Plus généralement, les technologies de l'information couplés avec les appareils connectés permettent de mieux gérer les ressources et services urbains, tels que l'énergie, la qualité de l'air et les parkings. La motivation majeure serait d'apporter une nette amélioration d'un point de vue environnemental, économique ou social. Ericson Mobility Report estime que d'ici 2024 il pourrait y avoir plus de 22 milliards d'objets connectés à travers le monde.

Les systèmes de télécommunications ont des besoins très différents en termes de portée, de volume et de fréquences d'informations échangées. Ainsi les objets connectés peuvent envoyer plus d'une centaine de messages par jour, à des débits d'environ 385 kilobits par seconde (comme par exemple pour le protocole LTE-M). Ceci multiplié par le nombre de dispositifs connectés sans cesse grandissant, pousse les experts du domaine à proposer des techniques de connectivité massive tout en garantissant des performances fiables ainsi qu'une meilleure qualité de service.

2 Résumé

Jusqu'à présent, les chercheurs et experts du domaine se sont surtout concentrés sur des approches d'accès multiples basées sur les protocoles RA (en anglais Random Access). Pour cause, cette technique, implémentée sur la couche MAC (Medium Access Control) du niveau 2 (Liaison de données) du modèle OSI, offre l'avantage de ne pas surcharger le réseau par le trafic de signalisation généré par les processus d'établissement de liaison, tel que le handshaking lors de la connexion initiale entre les appareils connectés et la station de base. Autre avantage majeur de ces protocoles, la station de base n'est plus responsable de l'allocation des ressources pour un trafic donné, car chaque utilisateur peut envoyer les données sans, au préalable, en avoir l'autorisation à condition de suivre la procédure prédéfinie. Ce qui augmente considérablement la limite du nombre d'objets connectés pouvant être supportés par le système.

Parmi les approches de type RA, nous pouvons citer les plus prometteuses telles que le NOMA (Non Orthogonal Multiple Access) qui superpose les signaux des appareils sur la même bande de fréquence, l'ACB (Class

Barring Access schemes) qui limite les congestions en interdisant temporairement (en fonction de paramètres prédéfinis) l'accès réseau à certains dispositifs connectés, le CSMA/CA qui évite les collisions en respectant trois principes : la fenêtre de contention, l'espace intertrame et les acquittement, et la dernière approche est Slotted-Aloha-base qui divise le temps de transmission d'un canal en intervalle de temps appelé "time slot". L'envoi de données n'est autorisé qu'au début d'un time slot. Si l'objet connecté n'effectue pas la transmissions à temps, il doit attendre le prochain time slot, ce qui réduit considérablement la probabilité de collision.

Les protocoles basés sur la technique RA sont moins complexes, néanmoins leurs performances restent, selon les auteurs, limité par leur caractère réactive, car elle ne font que réagir à la demande du trafic. Pour répondre à ces problème de performances, les auteurs proposent une toute nouvelle approche, dite proactive, basée sur la méthodologie de prédiction et ordonnancement conjoints (en anglais JSF pour Joint forecasting-scheduling). Le but de cette approche est de prévoir le modèle de génération de trafic de chaque dispositif connecté afin de préallouer des canaux de transmissions. De ce fait, le nombre de control signalant des congestions est réduit au minimum, les retards provoqués par les conflits, les collisions et les handshaking seront évités. Afin d'obtenir le meilleur Qos (Quality of service) possible, l'approche JSF se base sur l'algorithme MLP (Multi-Layer Perceptron) pour effectuer les prévisions. Il s'agit d'un algorithme de type réseau neuronal à propagation directe. Celui-ci combiné avec un système d'ordonnancement adapté peut considérablement accroître le débit attendu, minimiser les retards dans les transmissions et ainsi améliorer les performances du réseau.

Dans ce travail écrit nous nous intéresserons plus particulièrement à cette nouvelle approche conçu par les ingénieurs-chercheurs Volkan Rodoplu, Mert Nakip, Deniz Tursel Eliyi et Cunevt Guzelis. En effet, dans l'article intitulé "A Multi-Scale Algorithm for Joint Forecasting-Scheduling to Solve the Massive" les auteurs décrivent cet algorithme dont le but est de prévoir le trafic réseau d'un dispositif Iot en vue d'allouer en avance un canal de Liaison montante. En raison de sa nature multi-échelle (opérant à plusieurs échelle de temps), cet technique garantie une complexité temporelle et spatiale évolutive permettant de supporter plus de 6650 dispositifs iot. De plus, l'algorithme permet de maintenir le pourcentage du control de congestion du

débit à un niveau relativement bas, soit inférieur à 1,5 pourcent. Garantissant ainsi un haut débit de transmission. Le MSA est implémenté dans une passerelle Iot afin de répondre au problème d'accès massive survenant dans la couche MAC.

Des études ont démontrées que, sur plus de 3000 appareils connectés, les performances de cette approche était significativement plus élevées que trois autres protocoles conçus par les auteurs, qui proposent des approches tout aussi intéressante. Ainsi des comparaisons de consommation énergétique et de débit ont été effectués entre le MSA-JFS et les méthodes suivantes. Le premier protocole est le Reservation-based Access Barring (RAB) qui est une technique de type réactive combinant des fonctionnalités des approches ACB (Access Class Barring) et de RBS (Reservation Based Scheduling), le deuxième est le protocole de priorité basé sur la charge moyenne (en anglais Priority Based on Average - PAL) , il s'agit d'une méthode proactive se basant sur l'algorithme d'ordonnancement non préemptif qui classe les signaux par ordre de priorité en fonction de la charge de trafic restante, calculée sur base de la durée moyenne d'un délai de transmission. La troisième technique, Enhanced Predictive Version Burst (E-PRV-BO) est de type proactive, elle est issue du protocole PRV-PO (Predictive Version Packet Oriented) dont elle améliore certaines fonctionnalités.

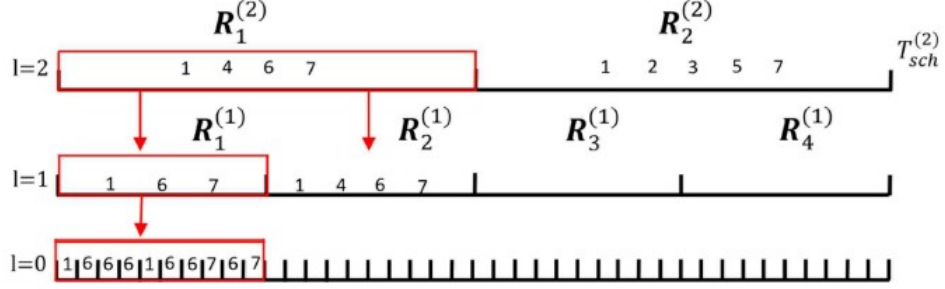
Dans ce système basé sur l'algorithme MSA, le gateway execute un forcaster, pendant un créneau temporelle spécifique, pour prévoir le modèle de trafic inhérent à chaque dispositif IOT dans le but de planifier un intervalle de temps dans lequel les liaisons montantes seront transmises sans risque de collision avec les liaisons descendantes. Pour ce faire, la technique de duplexage TDD (Time division Duplex) est utilisée. Celle-ci a la particularité de séparer les voies montantes et descendantes par un intervalle de garde, tout en garantissant que les caractéristiques du canal soient semblables sur les deux voies. Le gateway s'appuie sur le dernier modèle de trafic d'un dispositif connecté afin de prévoir le prochain modèle. Pour réaliser cette tâche, il garde constamment en mémoire les caractéristiques du modèle courant qu'il utilise dans le processus d'apprentissage du trafic.

Le fonctionnement des algorithmes MSF (Multi-scale Forecasting) est le suivant, pour chaque dispositif connecté désigné par i , une période de

génération de trafic est définis, τ_i . Elle correspond à la durée minimale entre chaque instance de génération de trafic successive. Le temps requis par les dispositifs connectés pour la génération d'un packet est un multiple de cette intervalle (en supposant que les dispositifs génèrent des données à intervalle régulier). Cette période est en corrélation avec l'horizon de prévision τ_i^{af} . Il s'agit du laps de temps durant lequel le gateway G effectue une prévision précise du trafic d'un dispositif connectés. Les prévisions peuvent être précises jusqu'à un certains nombre d'échantillons désigné par K_i . Soit pour chaque dispositif i , $\tau_i^{af} = K_i \tau_i$. Une des contraintes importantes de cette approche est que la plus longue intervalle de planification possible, désignée par τ_{sch} , doit être égale à l'horizon de prévision minimum de l'ensemble des appareils de la zone de couverture du gateway. Ainsi une planification des canaux de transmission ne pourra pas être effectuée correctement, car les appareils dont l'horizon de prévision est supérieur à cette limite ne seront pas pris en compte. L'approche novatrice dans ce travail est que l'ensemble des bits provenant d'une station est représenté sur plusieurs échelles de temps. Pour ce faire, les bits sont additionnées, récursivement, à des blocs de slots (des intervalles de temps) de plus en plus large. Aussi, nous avons la formule suivante : $X_i(B) = \sum_{b \in S_B} X_i(b)$. Au niveau 0 (le plus bas), la taille d'un bloc correspond à la durée de l'intervalle dans couche MAC. A chaque niveau l'ensemble des bits générés par une station, désigné par $X_i(b)$, arrive dans un bloc respectif, désigné par b . Ensuite à chaque niveau supérieur $\iota + 1$, les blocs sont regroupés en ensembles contingus de blocs. Ainsi, par exemple, le niveau ι contient l'ensembles des sous-blocs (S_B) qui formeront le bloc B dans le niveau $\iota + 1$. Nous pouvons donc définir $X_i(B)$, comme l'ensemble des bits d'une station au niveau $\iota + 1$.

Pour illustrer ce procédé, nous allons décrire le cas d'une opération à 3 niveaux, donc $L + 1 = 3$. Nous commençons le processus de planification au deuxième niveau $\tau_{sch}^{(2)}$ et définissons $R_b^{(\iota)}$ comme un sous-bloc b (slot de couche MAC) au niveau ι . Au lancement, une fonction déquillage de charge est exécutée, ce qui génère deux sous-blocs. Dans le premier $R_1^{(2)}$, se trouve les données des stations (1, 4, 6, 7) dont le nombre de bits de prévision est positive, idem pour le bloc $R_2^{(2)}$ qui selectionne les stations 1, 2, 3, 5, 7. Après ça, c'est au tour des blocs $R_1^{(1)}$ et $R_2^{(1)}$. Dans ce niveau seuls les stations du sous-bloc $R_1^{(1)}$ sont pris en compte pour la planification des ressources. Ce qui fait que ce bloc atteint le cas de base de la récursion et donc le MSA

effectue la planification des ressources des stations 1,6,7 en se basant sur la prévision des rafales au niveau 0. L'algorithme continuera le processus recursive jusqu'à ce que toutes les stations du niveau deux soient planifiées.



3 Conclusion

Bien que proposant des solutions concrètes pouvant répondre au problème d'accès massive dans une architecture qui utilise un Gateway IOT, les auteurs estiment qu'elle présente néanmoins quelques limites. Premièrement, vu que le gateway doit effectuer des prévisions et des planifications pour chaque station, et ce sur plusieurs échelles de temps, le temps de calcul est considérable lorsqu'il y a plus de 6650 stations. Deuxièmement, les simulations ne prennent pas en compte les transmissions multi-canaux, ni même le transfert d'une zone de couverture à une autre.