Demand response solutions Based on connected appliances

Rim Kaddah

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2016-ENST-0025

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité Informatique et Réseaux

présentée et soutenue publiquement par

Rim KADDAH

le 15 avril 2016

Gestion Active de la Demande Basée sur l'Habitat Connecté

Directeur de thèse: **M. Daniel KOFMAN**

Jury

M. György DÁN, Maître de Conférence, KTH, Suède	Rapporteur
M. Adam OUOROU, Chargé de Projet de Recherche, Orange Labs, France	Rapporteur
Mme Ana BUŠIĆ, Chargé de Recherche, Inria, France	Examinatrice
M. Gérard MEMMI, Professeur, Télécom ParisTech, France	Examinateur
M. David MENGA, Chercheur, EDF R&D , France	Examinateur
M. Michal PIÓRO, Professeur, Warsaw University of Technology, Pologne	Examinateur
Mme Maria POTOP-BUTUCARU, Professeur, UPMC, France	Examinatrice
M. Philippe FUTTERSACK, Chef de groupe, EDF R&D, France	Invité
M. Daniel KOFMAN, Professeur, Télécom ParisTech, France	Directeur de thèse
M. Fabien MATHIEU, Chercheur, Nokia Bell Labs, France	Encadrant

T H È S E

TELECOM ParisTech
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

PhD ParisTech

A dissertation presented

in fulfillment of the requirements for the degree of Doctor of Philosophy of

TELECOM ParisTech

Speciality Computer Networking and Telecommunications

presented and defended in public by

Rim KADDAH

On April 15 2016

Demand Response Solutions
Based on Connected Appliances

PhD Director: **Daniel KOFMAN**

À mes chers parents, mon frère et ma soeur.

Abstract

The equilibrium between generation and consumption is crucial for a proper functioning of the power system. The growing penetration of renewable intermittent energy sources, among other factors, is raising challenges to maintain this equilibrium. Demand Response (DR) contributes to overcoming these challenges by exploiting consumption flexibility in order to reduce system costs and increase reliability. . The Internet of Things (IoT) paradigm brings an opportunity for advanced Demand Response solutions. Indeed, it enables visibility and control on the various appliances that may consume, store or generate energy within a home.

In this thesis, we consider solutions having the capability to produce direct control decisions at different granularities based on variables measured at homes. We particularly focus on the optimal control of the appliances during time periods where the available capacity is not enough to satisfy the demand generated by homes. Control schemes are driven by an optimization based on utility functions. These functions are defined based on a generic approach that considers load's flexibility and the impact of control decisions on users. The proposed approach does not impose any restrictions on the type of controlled appliances nor on the granularity of control decisions. This enables joint control of heterogeneous loads. We consider three types of control architectures, namely centralized, partially distributed and fully distributed solutions. Schemes based on these architectures differ in the distribution of decision making among entities involved in the control and data that is made available to these entities. Numerical analysis shows the trade-offs of proposed solutions from a performance, scalability and complexity perspectives.

KEY-WORDS: Demand Response, Direct Load Control, Smart Grids, Internet of Things, Operations Research

Résumé

L'équilibre entre la production et la consommation est essentielle pour un bon fonctionnement du système électrique. Le déploiement croissant des sources d'énergie renouvelables intermittentes, parmi d'autres facteurs, rend cet équilibre plus difficile à maintenir. La Gestion Active de la Demande (GAD) est une solution capable de contribuer à répondre aux besoins d'équilibre grâce à l'exploitation de la flexibilité de la consommation. Cela permet de réduire les coûts du système et d'augmenter sa fiabilité. L'Internet des Objets (IdO) et le déploiement des équipements connectés permettent la mise en place de solutions avancées. En effet, il devient possible d'avoir plus de visibilité et un contrôle fin sur différents équipements qui consomment, stockent ou produisent de l'énergie dans une maison.

Dans cette thèse, nous considérons des solutions ayant la capacité de produire des décisions de contrôle direct à différents niveaux de granularité en fonction des variables mesurées dans les habitats. Nous nous sommes concentré particulièrement sur le contrôle optimal des équipements durant des périodes où la capacité de génération disponible ne suffit pas pour satisfaire la demande générée par toutes les maisons. Le contrôle est basé sur une optimisation d'utilité perçue. Des fonctions utilité sont définies à travers une approche générique qui considère la flexibilité de la charge et l'impact des décisions de contrôle sur les utilisateurs. L'approche proposée n'impose pas de restrictions sur le type des équipements contrôlés ni sur la granularité des décisions de contrôle. Ceci permet un contrôle joint d'équipements hétérogènes. Nous considérons trois types d'architectures de contrôle à savoir: des solutions centralisées, partiellement distribuées et entièrement distribuées. Ces architectures diffèrent dans la distribution de la prise de décision entre les entités impliquées dans le contrôle et les données qui sont mis à disposition de ces entités. L'analyse numérique montre les compromis des solutions proposées du point de vue de la performance, de l'extensibilité et de la complexité.

MOTS-CLEFS: Gestion active de la demande, contrôle direct des équipements, réseau électrique intelligent, Internet des Objets, recherche opérationnelle

Publications

[KKP14] Rim Kaddah, Daniel Kofman, and Michal Pioro. Advanced demand response solutions based on fine-grained load control. *2nd IEEE International Workshop on Intelligent Energy Systems (IWIES)*, pages 38–45, 2014.

[KKMP15] Rim Kaddah, Daniel Kofman, Fabien Mathieu, and Michal Pioro. Advanced demand response solutions for capacity markets. *11th International Conference on Innovations in Information Technology (IIT)*, 2015.

[MKBM16] Joel Mathias, Rim Kaddah, Ana Busic, and Sean Meyn. Smart fridge / dumb grid? demand dispatch for the power grid of 2020. *49th Hawaii International Conference on System Sciences (HICSS)*, 2016.

[KKMP16] Rim Kaddah, Daniel Kofman, Fabien Mathieu, and Michal Pioro. Semi-Distributed Demand Response Solutions for Smart Homes. Information Innovation Technology in Smart Cities, Springer Book.

[KM16] Rim Kaddah and Fabien Mathieu. Distributed demand response based on peer-to-peer communications. To be submitted.

Acknowledgements

I would like to thank Daniel Kofman for giving me the opportunity to work on a hot topic, for his time and guidance. I would like to thank Fabien Mathieu for his support, his time and his help. I would also like to thank him for his supervision towards the end of my PhD and his patience that were essential for completing my thesis.

I would like to thank Michal Pióro for his skills and time. I learned a lot from his extensive knowledge and experience in optimization. I would like to thank Ana Bušić, Sean Meyn and Joel Mathias for giving me the opportunity of working with them. They helped me develop my knowledge in control.

I would like to thank Marco for his time, his guidance and for the interesting discussions on optimization. I would also like to thank him, Fabien and Maura for reading my dissertation and for their relevant comments.

I would like to thank my office mates Jose, François, Julian, Şükrü and The Dang for the great moments and good humour. I would like to thank François, Fabien, Marco, Ludovic, Loïc, Alonso, Jose, Rémy for all the manipulation sessions. I would like to thank all my fellow labmates at LINCS with whom I shared great moments. In particular, I wish to thank Ahlem, Nivine, Sameh, Ghida, Riad, Jordan, Ahlem, Nihel, Nadjet, Stéphane, Michele, Natalia, Andrea, Sara, Yu-Ting, Yixi, Diana and Amy. I would like to thank my friends in Paris for the good moments, shared experience and advice particularly Mireille, Molly, Nour, Karam and Karim.

Mostly, I want to thank my family for their continued support all along: my mother, my father, my brother and my sister. I would like to thank my uncle Fahed for supporting me and helping me start the adventure in France.

Résumé en français

1 Introduction

Le réseau électrique est un système complexe en charge de relier les sources de génération et les consommateurs d'énergie. Un bon fonctionnement de ce système nécessite d'équilibrer la génération et la demande. Traditionnellement, l'équilibre est maintenu en contrôlant la génération : la production suit la courbe de charge. Cela implique de fortes contraintes sur le bouquet de production d'énergie comme le besoin d'un dimensionnement adéquat de la capacité du système pour pouvoir suivre des variations rapides de la demande. Aujourd'hui, ces contraintes sont satisfaites grâce à des types adaptés de sources de production : de haute puissance à réponse lente comme le nucléaire ; de faible puissance mais haute réactivité comme l'hydroélectrique ; intermédiaire comme les centrales à charbon. Certains de ces générateurs sont très coûteux en termes de maintenance et/ou d'impact environnemental.

Les tendances actuelles rendent cet équilibre précaire plus difficile à maintenir : d'une part, il n'y a pas de raison pour que la courbe de consommation devienne plus facile à suivre ; d'autre part, des objectifs environnementaux supplémentaires s'imposent comme le but d'atteindre 20% de géneration en provenance de sources d'énergie renouvelable à l'horizon 2020 en Europe.

Dans ce nouvel écosystème, la flexibilité et la surveillance fine sont des besoins clés qui permettront de réduire les coûts du système et d'optimiser son fonctionnement. En réponse à ces besoins et afin d'assurer la stabilité du réseau, le concept de *Smart Grid* a émergé. Smart Grid (ou réseau électrique intelligent en français) est une vision dans laquelle information, communication et contrôle sont réunis pour atteindre un système électrique durable, efficace et sécurisé [91].

Afin d'aider la génération à maintenir l'équilibre avec la demande, deux solutions sont envisageables dans un Smart Grid : le stockage d'énergie ou le contrôle de la demande. Le stockage d'énergie est la meilleure solution pour compenser les déséquilibres en complétant la production pendant les périodes de forte consommation ou en permettant une utilisation ultérieure de l'excès de production. Cependant, les technologies de stockage sont encore chères et non rentables (voir [66] et [19]).

Une solution alternative serait d'agir sur la demande. En effet, l'équilibre peut être maintenu en modifiant la courbe de la demande pour atteindre des objectifs de capacité de production liés à des limitations physiques ou financières. Les avantages du contrôle de la demande sont :

- sa grande réactivité comparé au contrôle de générateurs imposant des contraintes de démarrage. Ainsi, il est possible de réduire ou d'augmenter la demande instantanée en exploitant l'inertie de certaines charges (par exemple, le chauffage dans une maison peut être vu comme une batterie virtuelle).

- sa capacité à résoudre des contingences spatiales locales en raison de la distribution de la flexibilité sur le réseau (comparé à des générateurs raccordés de manière statique au réseau de transmission).

Les services conçus pour modifier la forme de la courbe de la demande en réponse aux exigences du réseau sont appelés services de la Gestion Active de la Demande (GAD). Deux types de solutions de GAD peuvent être envisagés pour atteindre les objectifs de forme de la demande : par incitation tarifaire ou par contrôle direct de la demande. Pour le premier type de GAD, le comportement des utilisateurs est influencé par un changement dynamique du prix de l'électricité au fil du temps. Pour le second, la consommation des usagers est directement contrôlée tout en leur offrant des paiements incitatifs pour accepter un tel contrôle intrusif. Le principal avantage du second type par rapport au premier est la garantie stricte du résultat à l'issue du contrôle.

Bien que les cas d'utilisation de la GAD deviennent plus aboutis ainsi que la technologie des composants permettant le pilotage de la charge [46], des solutions de GAD efficaces sont encore nécessaires. Ceci nécessite la définition d'une architecture de contrôle : l'interaction entre les entités en faisant partie ainsi que la distribution de l'intelligence entre ces entités doit être définie. Ensuite, en fonction de l'architecture spécifique, des algorithmes de contrôle doit être conçus. Ils doivent être capables de supporter des ressources hétérogènes pour fournir des degrés de liberté différents exploitables par le contrôle. Surtout, les solutions doivent être élaborées en vue de l'acceptation de l'usager. Répondre à ces besoins est l'objectif principal de cette thèse.

Comme les technologies sont encore en cours de développement pour fabriquer des batteries rentables, dans cette thèse, nous nous concentrons plutôt sur des solutions ciblant le pilotage de la charge par contrôle des équipements. Plus précisément, nous considérons des mécanismes basés sur le contrôle direct. Ce choix est motivé par la haute valeur d'un contrôle strict pour minimiser l'incertitude liée au comportement des usagers.

Pour concevoir des mécanismes de contrôle visant l'efficacité énergétique et la satisfaction des usagers, nous proposons une approche générique pour définir la flexibilité de la charge et l'impact des décisions de contrôle sur les usagers. Le framework proposé n'impose pas de restrictions sur le type des équipements contrôlés, ni sur la granularité des décisions de contrôle. Cela permettra un contrôle joint d'équipements hétérogènes.

Basé sur ce framework, nous instancions des problèmes de recherche opérationnelle. Ces problèmes représentent des mécanismes qui diffèrent dans les données mises à disposition et de la distribution de la prise de décision entre les entités impliquées dans le contrôle. En effet, nous examinons trois types d'architectures de contrôle, à savoir centralisées, partiellement distribuées et entièrement distribuées. Nous évaluons les limites de chaque solution proposée et nous discutons leur efficacité du point de vue des usagers en effectuant une analyse numérique sur des cas simples mais représentatifs d'utilisation. Nous analysons également l'évolutivité, la complexité et la sécurité (à savoir, la confidentialité et l'anonymat).

Les contributions de cette thèse sont présentées en commençant par un résumé de l'état de l'art du Chapitre 2 du mémoire dans la Section 1. Ensuite, dans la Section 1, nous présentons le *framework* (cadre) que nous proposons pour décrire un équipement générique (Chapitre 3 du mémoire). Ce framework est utilisé pour définir une méthode de construction de fonctions d'utilité capable de prendre en compte la nature de l'équipement contrôlé ainsi que sa priorité. Section 1 présente quelques mécanismes de contrôle proposés qui sont sélectionnés de manière à montrer les conclusions essentielles de cette thèse (Chapitres 4, 5 et 6 du mémoire). Une analyse numérique sur un cas d'usage spécifique est présentée dans la Section 1. Finalement, la Section 1 présente les conclusions et perspectives correspondant au Chapitre 8 du mémoire de thèse.

2 État de l'art

Bien qu'un nombre important de services de GAD pour des utilisateurs résidentiels sont proposés, très peu sont ceux qui se sont intéressés aux systèmes de contrôle où différents types d'équipements sont modélisés. En effet, la plupart se sont focalisés sur des équipements spécifiques flexibles (par exemple, les piscines, chauffage, climatisation).

Les solutions dans [97, 55, 89] sont des exemples de propositions qui prennent en compte de la flexibilité d'un équipement générique.

Pour cela, les auteurs de [55] proposent de classer les équipements en quatre types : le premier type est constitué d'équipements qui contrôlent la température dans la maison, le second type comprend les équipements qui ont besoin de compléter leur tâche avant une date limite, le troisième type comprend les équipements qui ont un effet immédiat sur les usagers (par exemple, l'éclairage), et le quatrième type comprend des équipements qui ne sont pas critiques et pour lesquels les usagers sont préoccupés par leur consommation d'énergie (par exemple, les équipements de divertissement). Leur modèle est utilisé pour proposer une solution de contrôle basée sur une tarification dynamique de l'énergie et ne prend pas en compte l'équité ou la différenciation de service entre usagers. La différenciation de service peut être introduite par la notion de priorité comme la proposition de [50] qui considère différents niveaux de priorités à traiter séparément, et que nous étendrons pour traiter l'équité pour un niveau de priorité donné.

Les auteurs dans [97] proposent un système de contrôle temps réel qui récompense les usagers pour accepter le contrôle de leurs équipements. Les usagers doivent fournir toutes les informations liées aux équipements contrôlés pour évaluer leur importance et la flexibilité qu'ils peuvent offrir en fonction des contraintes définies par l'usager. La proposition de [89] est la plus proche de notre approche de modélisation et notre système de contrôle hiérarchique. Cependant, dans leur approche, l'équité n'est pas prise en compte. En plus, leur système est itératif et requière une convergence pour obtenir une solution qui ne viole pas les contraintes du problème liées à la capacité limitée du système.

Les solutions exposées précédemment donnent une importance significative à l'agrégateur (une entité centrale responsable de produire les décisions au niveau global du système). Ceci peut causer des problèmes de point de défaillance unique qui peuvent être atténués lorsque des systèmes de contrôle distribués sont considérés. En effet, ces systèmes assurent une plus grande robustesse et évolutivité en limitant l'échange d'informations avec le niveau d'agrégation. La plupart des systèmes distribués sont proposés dans le contexte des services de GAD basés sur la tarification de l'énergie et la recharge de véhicules électriques. Dans [85], les auteurs proposent un mécanisme de GAD qui définit des prix dynamiques en deux étapes et qui est très proche de notre approche de contrôle distribué. Dans la première étape, un système hiérarchique est supposé dans lequel le fournisseur d'énergie (l'agrégateur) annonce des prix qui sont plus élevés lorsque la consommation dévie plus loin de sa valeur moyenne. Sur la base des prix, les usagers planifient leurs équipements et sauvegardent leur coût minimal réalisable. Dans la deuxième étape, les usagers peuvent se coordonner pour aplanir leur consommation d'énergie. La coordination est validée si deux usagers arrivent à avoir un coût inchangé ou plus bas. Les auteurs montrent que, en étant sincère, les usagers peuvent réaliser la plus grande diminution du coût. Le système de coordination proposé exige de propager la charge globale entre les usagers ce qui empêche une parallélisation des échanges. Dans notre proposition, nous présentons une approche qui permet aux usagers de coordonner avec d'autres usagers à tout moment. En outre, nous supposons des contraintes strictes sur l'énergie qui peut être consommée par les usagers comme nous nous intéresserons à une situation où la capacité du système n'est pas suffisante pour satisfaire tous les demandes.

3 Modèle

Dans cette section, nous proposons un framework générique et un modèle du système sur lequel nous nous basons pour définir les architectures de contrôles présentées dans les sections suivantes. En effet, nous allons nous concentrer sur la présentation de l'architecture générale que nous utilisons pour concevoir les mécanismes de contrôle direct visant une réduction de la consommation en-dessous d'une limite de capacité souhaitée durant une période de temps cible. Afin de permettre un contrôle joint d'équipements hétérogènes, nous proposons une taxonomie des équipements capable de décrire leur flexibilité. Sur la base de cette taxonomie, nous proposons une approche pour définir des fonctions d'utilité permettant de guider le contrôle et d'évaluer son impact sur la qualité d'expérience perçue par l'usager.

3.1 Architecture

Étant donné que le réseau de distribution électrique est hiérarchique, nous proposons une architecture de contrôle ayant une structure similaire. Cela permet de prendre en compte d'éventuelles contraintes physiques à différents points d'agrégation du réseau.

Nous considérons un modèle simple à deux niveaux composés chacun d'entités appartenant à un groupe fonctionnel : un côté de agrégateur ou l'opérateur du réseau de distribution et un côté «maisons». Nous les appelons entité de décision de l'agrégateur (DE_a) et entité de décision de la maison (DE_h) respectivement. Cette architecture peut être généralisée en considérant une architecture à niveaux multiples. Cependant, nous considérons qu'un modèle simplifié est suffisant pour capturer les problèmes essentiels qui se posent lorsque différents niveaux de décisions sont prises en compte. Nous discutons de l'architecture multi-niveaux pour les systèmes de contrôle pour lesquels c'est pertinent.

Pour une maison donnée, la DE_h est au courant de l'état de certaines variables surveillées chez l'usager et reçoit des ordres de contrôle de la DE_a. Selon le contrôle spécifique considéré, la DE_h peut transmettre tout ou une partie des données collectées à la DE_a. Nous définissons trois familles de solutions de GAD qui dépendent des spécificités de la communication entre les différentes entités du système :

- *Les solutions centralisées :* la DE_a contrôle directement les équipements dans les maisons (qui représentent les feuilles de la hiérarchie). Pour produire ses décisions, elle doit avoir des informations statiques ou dynamiques complètes sur les équipements contrôlés et les variables pertinentes des maisons. Pour cette famille, le rôle de la DE_h est de relayer les signaux de contrôle aux équipements.

- *Les solutions partiellement distribuées :* la DE_a prend des décisions à la granularité des DE_h. Les décisions peuvent être basées sur des informations agrégées statiques ou dynamiques. Ainsi, la DE_a ne peut pas avoir une connaissance directe des informations liées aux équipements individuels ou de l'état des variables surveillées dans les maisons. Dans ce cas, le contrôle des équipements est effectué par les DE_h.

- *Les solutions distribuées :* les DE_h interagissent pour améliorer la qualité d'expérience perçue dans le système. Ces solutions considèrent un état initial du système dans lequel les DE_h ont une certaine visibilité sur les contraintes du système (capacité de consommation). Le contrôle des équipements est toujours effectué par les DE_h de chaque maison.

Ces solutions sont proposées et évaluées dans le contexte d'une période de contrôle finie et connue durant laquelle le réseau manque de capacité de génération et ne peut pas répondre à toutes les demandes générées par les équipements contrôlés.

3.2 Taxonomie

Afin de développer des mécanismes de contrôle, nous proposons de définir un framework générique permettant une séparation entre l'impact de l'usage d'un équipement et les contraintes liées au contrôle. Nous proposons donc de caractériser les équipements par les trois attributs suivants :

Attribut 1 - Usage

Cet attribut exprime les attentes liées à l'usage d'un équipement. Il est décidé par l'usager en cas d'ambiguïté en fonction du type d'usage destiné. Inspiré par [55], nous proposons trois types d'usage :

- *Les demandes interactives* sont celles qui sont déclenchées directement par les usagers et, si possible, doivent être servies immédiatement (similaire au type 3 dans [55]). Retarder le service de ces demandes n'a pas de sens. L'usage d'équipements relevant de cette catégorie est souvent directement dépendant de la présence d'habitants. Les système d'éclairage et de divertissement (par exemple, TV) sont des exemples typiques de demandes interactives.

- *Les charges de fond* sont celles pour lesquels les usagers expriment une préférence sur la valeur de certaines variables de la maison affectées par l'opération de l'équipement en question et non pas directement sur le profil de puissance (similaire au type 1 dans [55]). L'opération du chauffage est un exemple de charge de fond pour laquelle l'usager exprime une préférence de température de la pièce. Les charges de fond peuvent être souvent avancées ou retardées sans impacter significativement la qualité d'expérience de l'usager.

- *Les charges basées sur un programme* sont celles qui ont des cycles de fonctionnement programmables. Cet usage est similaire au type 2 dans [55]. La valeur de l'opération d'un équipement ayant ce type d'usage est liée à la capacité de compléter un programme. La flexibilité principale fournie est liée à la possibilité d'opérer l'équipement à tout moment entre une date de début et une date de fin à laquelle le programme doit être complété. Un exemple typique d'équipement appartenant à cette catégorie est une machine à laver.

Attribut 2 - Criticité et préférences

Indépendamment de l'usage, certains équipements sont plus critiques que d'autres ([50]) : par exemple, un équipement médical à domicile doit fonctionner en cas de besoin et est plus critique qu'un équipement de divertissement. Cette criticité peut être subjective définie par l'usager ou même dépendre de l'état de l'équipement et de la maison. Dans tous les cas, les décisions de contrôle doivent pouvoir prendre en compte cette criticité afin de répondre aux besoins plus critiques avant d'essayer d'atteindre les objectifs de criticité inférieure. Dans notre proposition, nous supposons différents niveaux de criticité. Chaque équipement commandé peut remplir un ou plusieurs besoins correspondant aux différents niveaux. Dans cette thèse, nous considérons deux niveaux de criticité que nous nommons *vital* et *confort*. Le premier exprime des besoins hautement prioritaires (par exemple, un certain niveau d'éclairage, le chauffage quand la température est très basse) et le second exprime des préférences moins essentielles (par exemple, le chauffage quand la température est acceptable, l'éclairage décoratif). En plus de la criticité, les utilisateurs peuvent avoir des préférences de priorité entre les équipements servant le même groupe de criticité.

Attribut 3 - Caractéristiques électriques et de fonctionnement

Cette catégorie permet de distinguer les appareils en fonction de la façon dont ils consomment de l'énergie dans le temps. Les contraintes peuvent être imposées par la technologie spécifique de l'équipement (par exemple, résistive, inductive, stockage). Pour un équipement donné, nous distinguons deux types de flexibilité : une liée à la puissance et une liée au temps.

Une flexibilité liée à la puissance signifie qu'il est possible de modifier/réduire la puissance consommée par un équipement tout en lui permettant de fonctionner sans nuire à sa durée de vie. Un autre cas de flexibilité de ce type est d'agréger des équipements servant le même besoin en un gros «méta-équipement». Un exemple d'une telle simplification peut être proposé pour l'éclairage. En effet, la fourniture d'une certaine puissance pour l'éclairage permettra d'allumer un certain nombre d'ampoules.

La flexibilité liée au temps exprime les contraintes de durée minimale de fonctionnement requise. Ces contraintes peuvent être définies pour protéger la durée de vie de l'équipement ou pour interdire son interruption avant la fin de son cycle d'opération (par exemple, machine à laver).

3.3 Fonction d'utilité

Nous nous servons de la taxonomie introduite précédemment pour définir des fonctions d'utilité pour chaque équipement contrôlé et les contraintes de contrôle correspondantes. Ceci est réalisé en trois étapes :

1. Dériver les fonctions d'utilité individuelles pour chaque équipement pris seul (basé sur l'Attribut 1)

2. Introduire une priorité entre plusieurs équipements commandés (basé sur l'Attribut 2)

3. Définir les contraintes sur les décisions de contrôle pour chaque appareil (basé sur l'Attribut 3)

Nous détaillons maintenant chacune de ces étapes.

Fonctions d'utilité individuelles

Le contrôle des équipements doit prendre en compte l'impact de la décision sur l'usager. Pour guider les décisions, nous voulons concevoir un framework générique indépendant de la nature des équipements contrôlés et suffisamment flexible pour modéliser tout équipement. Nous visons également à rendre le processus automatique tout en permettant aux utilisateurs de définir manuellement leurs préférences subjectives.

Pour atteindre ces objectifs, nous définissons une fonction d'utilité individuelle $f_{u_l ht}^{a}$ pour chaque appareil a à la maison h au temps t pour le niveau de criticité u_l (introduit dans la Section 1). Cette fonction évalue la qualité instantanée de l'expérience telle qu'elle est perçue par l'utilisateur, sur la base de l'état de certaines grandeurs mesurables dans la maison, que nous appelons *variables contrôlables*. Une variable contrôlable peut être la consommation d'énergie ou une variable exogène comme la température. La valeur de cette fonction est définie entre 0 (à savoir aucune utilité perçue) et 1 (utilité maximale perçue).

Ainsi, la fonction d'utilité individuelle est définie comme suit :

Definition 1. *Soit m_a le nombre de variables contrôlables pertinentes à l'appareil a. Nous appelons une fonction d'utilité individuelle de l'appareil a au temps t dans la maison h pour le niveau de criticité u_l, une fonction $f_{u_l ht}^{a} : \mathbb{R}^{m_a} \rightarrow [0,1]$.*

Les variables contrôlables pertinentes à un appareil peuvent être déterminées grâce à l'usage de l'équipement (attribut 1). Des exemples de modélisation sont fournis dans la Section 1.

À ce stade, les équipements ont la même grandeur d'utilité. Cependant, certains appareils sont plus importants que d'autres. Le modèle d'utilité doit pouvoir exprimer une priorité entre des équipements différents dans une même maison ou dans des maisons différentes.

Gestion de priorité

Afin d'étendre le modèle pour prendre des décisions liées à des équipements dans la même maison ou dans des maisons différentes, nous avons besoin d'introduire la priorité et la criticité (Attribut 2 dans Section 1). Nous regardons maintenant comment faire évoluer le modèle de fonction d'utilité pour répondre à ce besoin.

Priorité

Pour définir la priorité entre les appareils ayant le même niveau de criticité u_l, nous devons d'abord prendre en compte l'échelle du contrôle. Concrètement, la priorité doit être définie entre les équipements d'une même maison et entre les maisons. Ceci est réalisé en mettant à l'échelle les fonctions d'utilité individuelles $f_{u_l ht}^{a}$ d'un appareil a à la maison h au temps t(voir la Section 1) pour obtenir les fonctions d'utilité réelles $U_{u_l ht}^{a}$ évaluées lors du contrôle.

Pour cela, dans chaque maison h, les résidents doivent définir pour chaque équipement a un score de priorité $\Phi_{u_l}^{a}(h)$. Ces scores sont utilisés pour pondérer les fonctions d'utilité individuelles afin de prioriser les équipements dans la maison.

Pour permettre aux maisons d'avoir différentes exigences de qualité de service, une utilité maximale $U_{u_l max}(h)$ est définie pour chaque maison h. Ainsi, une valeur élevée de $U_{u_l max}(h)$ correspond à un contrat «premium» alors que les maisons ayant une utilité maximale faible auront une qualité de services plus dégradée.

Supposons que la période de contrôle est discrétisée en T_m instants et A est le nombre total d'équipements commandés. En se basant sur l'utilité $f_{u_l ht}^{a}$ et le score $\Phi_{u_l}^{a}(h)$, la fonction d'utilité instantanée $U_{u_l ht}^{a}$ de l'équipement a à la maison h au temps t peut être exprimée par :

$$U_{u_l ht}^{a} = c_{u_l h}^{a}\, f_{u_l ht}^{a}$$

où $c_{u_l h}^{a} = \frac{\phi_{u_l}^{a}(h)U_{u_l max}(h)}{max(\sum_{t=1}^{t_M}\sum_{a'=1}^{A}\phi_{u_l}^{a'}(h)\ f_{u_l ht}^{a'})}$ est le facteur d'utilité pour l'appareil a à la maison h. Ce facteur d'utilité nous permet d'imposer une normalisation qui tient compte de l'utilité maximale réalisable pour chaque maison. En effet, si tous les f_{ht}^{a} peuvent atteindre 1, $c_{u_l h}^{a}$ peut être simplifié à $\frac{\phi_{u_l}^{a}(h)U_{u_l max}(h)}{t_M\sum_{a'=1}^{A}\phi_{u_l}^{a'}(h)}$. De cette manière, l'utilité maximale est automatiquement distribuée sur les équipements selon leur score de priorité. Lorsque les utilités individuelles atteignent leur valeur maximale à la maison h, l'utilité réelle totale sur la période de contrôle vaut $U_{u_l max}(h)$.

Dans ce qui suit, nous utiliseront les fonctions réelles $U_{u_l ht}^{a}$ dans la formulation des algorithmes de contrôle ciblant un niveau de criticité u_l. Nous discutons maintenant comment nous traitons la criticité dans le modèle d'utilité.

Criticité

Pour introduire la criticité, nous supposons un modèle de décision à plusieurs niveaux. En effet, la séparation des décisions est essentielle lorsque les niveaux de criticité sont différents. En fait, sans séparation, les politiques de contrôle peuvent considérer, par exemple, que l'éclairage d'un grand nombre d'ampoules est équivalent en termes d'utilité à la mise sous tension d'un appareil critique (par exemple une machine de dialyse). En outre, nous supposons qu'un équipement peut servir des niveaux de criticité différents. Nous définissons donc un tuple de valeurs d'utilité réelles correspondant chacun à un niveau de criticité. Supposons L niveaux de criticité et une utilité réelle notée U_{ht}^{a} pour l'équipement a au temps t dans la maison h. Nous pouvons écrire $U_{ht}^{a} = (U_{u_1 ht}^{a}, U_{u_2 ht}^{a}, .., U_{u_L ht}^{a})$ où u_1, u_2, .., u_L expriment des niveaux décroissants de criticité .

Lors du contrôle d'un ensemble d'équipements, les politiques visent à satisfaire les besoins les plus importants exprimés par le niveau u_1 avant d'améliorer d'autres niveaux d'utilité. Le même s'applique aux autres niveaux où chaque niveau doit être entièrement optimisé avant d'améliorer les valeurs d'utilité des niveaux inférieurs.

Comme indiqué dans la Section 1, nous supposons un modèle simplifié à deux niveaux d'utilité, nommés *vital* et *confort*. Ainsi, nous pouvons exprimer les utilités par des paires vital/confort : $U_{ht}^{a} = (U_{v}{}_{ht}^{a}, U_{c}{}_{ht}^{a})$. Les décisions de contrôle sont basées sur une comparaison lexicographique des valeurs d'utilité. Donc, pour deux valeurs d'utilité U_{ht}^{a} et $U_{ht}'^{a}$, nous considérons :

$$U_{ht}^{a} > U_{ht}'^{a} \text{ ssi } U_{v}{}_{ht}^{a} > U'_{v}{}_{ht}^{a} \text{ ou } (U_{v}{}_{ht}^{a} = U'_{v}{}_{ht}^{a} \text{ et } U_{c}{}_{ht}^{a} > U'_{c}{}_{ht}^{a}).$$

Les utilités peuvent être additionnées en les sommant élément par élément.

Type de contrôle

Le problème d'optimisation que nous visons à résoudre, consiste à décider de la quantité de puissance attribuée à chaque équipement. La qualité de la décision est mesurée par la valeur de la fonction d'utilité réelle définie précédemment.

Nous allons nous baser sur l'attribut 3 de la Section 1 pour identifier le type de contrôle qui peut être appliqué. Le contrôle peut être discret (ON-OFF) ou continu (puissance fournie). Des contraintes temporelles peuvent aussi être considérées pour modéliser la tolérance aux interruptions (par exemple, une fois que l'appareil est utilisé, il doit rester allumé pendant une certaine durée).

Pour tout équipement a dans une maison h, les contraintes de consommation d'énergie à un certain temps t peuvent être formulées en utilisant deux paramètres et deux variables de décision. Les paramètres sont une valeur minimale de puissance $P_m^a(h)$ et une valeur maximale $P_m^a(h)$ en watts. Les variables de décision sont la puissance X_{ht}^a allouée à l'appareil a à la maison h au temps t et une variable binaire x_{ht}^a (égale à 1 lorsque l'équipement est en marche et à zéro quand il est éteint).

Cette formulation est exprimée par l'équation (1).

$$P_m^a(h)x_{ht}^a \leq X_{ht}^a \leq P_M^a(h)x_{ht}^a. \tag{1}$$

Des équations supplémentaires peuvent être ajoutées à cette formulation pour exprimer des contraintes temporelles comme ceux liées à la tolérance aux interruptions .

4 Systèmes de contrôle proposés

Dans la section précédente, nous avons proposé un framework générique qui permet de définir des fonctions d'utilité pour tout équipement. Nous avons montré comment ce framework peut être utilisé pour prendre en compte la criticité et la priorité entre différents équipements. Maintenant, nous nous intéressons à la définition d'algorithmes de contrôle visant à réduire la consommation en dessous d'une limite désirée en maximisant l'utilité perçue par les usagers. Comme indiqué précédemment, nous allons nous concentrer sur trois familles de solutions de contrôle : centralisée, partiellement distribuée et distribuée.

4.1 Solutions centralisées

Une solution centralisée basée sur des informations à granularité fine est une solution envisageable pour résoudre le problème d'ordonnancement sous contraintes d'insuffisance de la capacité disponible. Dans cette thèse, nous avons proposé plusieurs algorithmes centralisés. Nous avons considéré différentes manière de définir l'équité et nous avons étudié des algorithmes de contrôle qui agissent à différentes échelles de temps, à savoir en avance ou en temps réel. Nous allons maintenant présenter certains des algorithmes proposés en supposant un niveau d'équité garanti à travers le modèle de fonction d'utilité à deux niveaux (vital et confort). En effet, dans ce qui suit, nous présenterons deux approches de contrôle : une qui agit en avance tout en visant une solution exacte au problème de maximisation d'utilité, et une qui agit en temps réel en approximant l'utilité de chaque appareil à tout instant. Nous supposons une discrétisation de la période de contrôle en t_M tranches de temps de durée constante. Nous supposons aussi le contrôle de A classes d'équipements dans H maisons participantes à la GAD.

Problème global

Nous supposons un contrôle exact où la DE_a a une vue complète sur le système et produit ses décisions en avance de manière à maximiser l'utilité totale du système sur la période de contrôle. Nous appelons cette solution «maximum global» et nous la notons par *GM* (pour *Global Maximum*).

$$\max_{X_{ht}^a, x_{ht}^a} \sum_{t=1}^{t_M} \sum_{h=1}^{H} \sum_{a=1}^{A} U_{ht}^a \tag{2a}$$

s.t.

$$\sum_{h=1}^{H} \sum_{a=1}^{A} X_{ht}^a \leq C(t), \ \ \forall\, t \tag{2b}$$

$$P_m^a(h) x_{ht}^a \leq X_{ht}^a \leq P_M^a(h) x_{ht}^a, \ \ \forall\, t, \forall\, h, \forall\, a \tag{2c}$$

$$x_{ht}^a \in \{0, 1\}, \ \ \forall\, t, \forall\, h, \forall\, a. \tag{2d}$$

Le problème (2) peut être résolu par la DE_a si toutes les informations sur les équipements et leurs fonctions d'utilité sont transmises par chaque DE_h. La DE_a peut alors calculer une solution optimale et aviser les maisons en conséquence.

Les variables de décision dans ce cas sont x_{ht}^a et X_{ht}^a (cf. 1).

Cette formulation n'explicite pas les contraintes liées à des équipements spécifiques (e.g. contraintes temporelles pour la machine à laver).

Algorithme glouton

Le problème présenté précédemment peut être résolu de manière exacte en utilisant un solveur de Programmation Linéaire Mixte en Nombres Entiers (PLMNE) tel que CPLEX [45]. Cette approche est coûteuse vu la NP-complétude du problème à résoudre. Pour pallier à cette complexité, nous proposons et évaluons des algorithmes gloutons. Dans cette section, nous allons nous focaliser sur un algorithme particulier vu sa performance en terme d'utilité globale comme nous le verrons dans la Section 1. D'autres variantes sont étudiées dans la version anglaise complète de la thèse (Chapitre 6.2).

L'algorithme

L'algorithme que nous présentons maintenant agit en temps réel. Cependant, les décisions sont prises en vu de la dynamique des variables du système ainsi que la connaissance des paramètres dans les tranches de temps futures (par exemple, la capacité disponible).

Nous considérons une discrétisation de la puissance pour réduire les options d'allocation possibles. En effet, comme la puissance allouée à un équipement peut prendre une valeur entre un minimum de puissance nécessaire pour faire fonctionner l'équipement et une puissance maximale, nous introduisons l'incrément ΔP. Cet incrément définit l'augmentation de puissance minimale des allocations.

Les décisions sont prises pour chaque tranche de temps de manière consécutive. Une liste L_2 est définie et remplie avec des éléments qui peuvent être alloués à une tranche de temps donnée. Informellement, un élément représente la décision d'allouer une puissance supplémentaire à un équipement donné. Pour cette tranche de temps, l'élément le plus rentable (gain d'utilité rapporté à l'augmentation de puissance) est choisi et l'équipement correspondant reçoit la nouvelle allocation de puissance. Après chaque décision, la liste L_2 est mise à jour et le processus est répété jusqu'à ce qu'aucune allocation ne puisse être réalisée au vu de la capacité disponible du système et des contraintes de fonctionnement de l'équipement à la tranche de temps supposée. Une fois le processus d'allocation sur une tranche de temps terminé, nous passons à la prochaine tranche de temps pour laquelle le même processus est appliqué. Ceci est fait pour tous les tranches de temps de la période d'optimisation.

Cette heuristique gloutonne temps réel est décrite par l'algorithme 1 et est désignée par RTGreedy.

Algorithm 1 Algorithme glouton temps réel

---

procedure RTGREEDY
  Étape 0 :
  Initialiser les variables $X^a_{ht} = 0, x^a_{ht} = 0,\ \ \forall\ t,\ \forall\ h,\ \forall\ a$;
  for $t = 1$ to t_M **do**
    Initialiser le poids de chaque élément $w^a_{ht} = P^a_m(h)$ et sa valeur v^a_{ht}, $\forall\ h,\ \forall\ a$;
    Initialiser la liste L_2 avec les éléments d'équipements qui peuvent fonctionner selon la capacité du système $C(t)$.
    while $L_2 \neq \emptyset$ **do**
      Étape 1 : Choisir l'élément le plus rentable dans L_2 (par exemple l'équipement a_i dans la maison h_j ayant le rapport $v^{a_i}_{h_jt}/w^{a_i}_{h_jt}$ le plus élevé).
      Étape 2 : Attribuer la puissance $w^{a_i}_{h_jt}$ à l'équipement a_i dans la maison h_j en mettant à jour les variables $X^{a_i}_{h_jt_k}$ et $x^{a_i}_{h_jt_k}$ pour $t_k \geq t$.
      Étape 3 : Mettre à jour la liste L_2 en modifiant le poids et la valeur des éléments correspondant à des équipements dont la puissance allouée peut être incrémentée selon la capacité restante $(C(t) - \sum_{h=1}^{H} \sum_{a=1}^{A} X^a_{ht})$.
  Sortie Puissances allouées X^a_{ht} et x^a_{ht}, $\forall\ t,\ \forall\ h,\ \forall\ a$.

Spécificités par type

Nous discutons maintenant comment cet algorithme traite chaque type d'usage pour obtenir la valeur v^a_{ht} d'une allocation de puissance w^a_{ht} à un équipement a dans une maison h. Quel que soit son type, chaque appareil a est représenté par $1 + \lceil (P^a_M(h) - P^a_m(h))/\Delta P \rceil$ éléments qui peuvent être alloués à chaque instant. Le premier élément alloué correspond à la puissance minimale requise pour faire fonctionner l'équipement.

Les demandes interactives

Pour les demandes interactives, les fonctions d'utilité sur les différentes tranches de temps sont généralement indépendantes. Pour cela, le problème d'allocation peut être considéré comme un problème de sac à dos unidimensionnel pour chaque tranche de temps. Ainsi, une décision optimale sur chaque tranche de temps est la décision optimale du problème global considérant tout l'intervalle d'optimisation.

La valeur de chaque élément dépend de la puissance déjà allouée et peut être directement dérivée de la fonction d'utilité pour une tranche de temps donnée. La valeur v^a_{ht} d'un élément d'un équipement interactif a est égal à la différence entre la valeur correspondant à son allocation $(X^a_{ht} + w^a_{ht})$ et la valeur correspondant aux éléments qui lui sont déjà alloués (X^a_{ht}). Ces valeurs correspondent à la pente de l'enveloppe convexe de la fonction d'utilité U^a_{ht}. Ceci permet aux éléments d'avoir une valeur décroissante quand la puissance allouée augmente. Ainsi, les éléments seront considérés dans le bon ordre comme le premier élément est plus efficace que les suivants et ainsi de suite.

Les charges de fond

Pour les charges de fond, lorsque l'impact des variables couvre plusieurs tranches de temps, la dépendance temporelle entre les décisions doit être prise en compte. Dans ce qui suit, nous élaborons sur le sujet en considérant l'exemple du chauffage. Le chauffage durant une tranche de temps t va générer une incrémentation de la température qui affectera les tranches de temps suivantes (en raison de l'inertie). Ainsi, la valeur du chauffage à une tranche de temps t est égale à la valeur d'utilité générée par le chauffage à t et les valeurs d'utilité acquises grâce à l'augmentation de la température dans les tranches de temps ultérieures due au chauffage à la tranche t.
La valeur du fonctionnement du chauffage a à une tranche de temps t peut être écrit : $v^a_{ht} = \sum_{t'=t}^{t_M} \Delta U^a_{ht'}$ où $\Delta U^a_{ht'}$ désigne le gain supplémentaire d'utilité en raison de l'augmentation de la température à la tranche de temps t' générée par le chauffage à t.
Nous considérons un modèle discret d'évolution de la température qui suit l'équation :
$T_{ht} = T_{h(t-1)} + F(h)(X^a_{ht} + w^a_{ht}) + G(h)(T_e(t) - T_{h(t-1)})$.

Appelons $\Delta T_{ht'}$ le changement de température à t' en raison du chauffage à $t \leq t'$. Ainsi, si le chauffage est augmenté en puissance à la tranche de temps t ($X^a_{ht} + w^a_{ht}$), un gain de température de $\Delta T_{ht} = F(h)w^a_{ht}$ est perçu à t. De proche en proche, nous mesurons le gain de température sur les tranches de temps suivantes causé par l'augmentation de l'allocation de w^a_{ht} en ne prenant à chaque fois que les termes de l'équation qui dépendent de la température précédente :
$\Delta T_{h(t+1)} = \Delta T_{ht}(1 - G(h))$;
$\Delta T_{h(t+2)} = \Delta T_{h(t+1)}(1 - G(h))$ et ainsi de suite.
Sur la base de ces équations, $\Delta T_{ht'}$ (pour $t' = t+1$ à t_M) peut être exprimé en fonction de ΔT_{ht} par l'équation :
$\Delta T_{ht'} = \Delta T_{ht}(1 - G(h))^{(t'-t)}$. Nous pouvons donc calculer la somme des gains en température :
$\sum_{t'=t+1}^{t_M} \Delta T_{ht'} = \Delta T_{ht}(1 - G(h))\frac{1-(1-G(h))^{(t_M-t)}}{G(h)}$.

Ce gain total va nous permettre de calculer la valeur du chauffage v^a_{ht} si la puissance w^a_{ht} est ajoutée. Cette valeur est estimée en prenant en compte le niveau de service qui va être rempli à la tranche de temps où l'allocation a lieu. Dans le cas de deux niveaux d'utilité (vital et confort), cela dépendra de la température étant au-dessus ou en-dessous de la température minimale tolérée $T_m(h)$. En effet, si la température est supérieure à $T_m(h)$, seule l'utilité de confort est supposée pour les température gagnées grâce au chauffage dans les tranches de temps suivants. Si la température est inférieure à $T_m(h)$, l'utilité vitale est supposée. En effet, même si la température est au-dessus de $T_m(h)$ à la tranche de temps où le chauffage est augmente en puissance, elle pourra descendre en dessous de $T_m(h)$ dans les tranches de temps futures. C'est une simplification du calcul exacte de l'utilité qui s'est avéré plus efficace dans nos analyses numériques.

Ainsi, pour calculer la valeur de chauffage v^a_{ht} de w^a_{ht}, nous prenons la valeur ΔU^a_{ht} correspondant à l'augmentation de $\Delta T_{ht} = F(h)w^a_{ht}$ de la température. Sur la base de la valeur ajoutée (vitale ou confort) au temps t, nous calculons $\sum_{t'=t+1}^{t_M} \Delta U^a_{ht'}$ en remplaçant $\sum_{t'=t+1}^{t_M} \Delta T_{ht'}$ par l'expression obtenue précédemment (fonction de ΔT_{ht}).

Les charges basées sur un programme

Comme les charges basées sur un programme ont des fonctions d'utilité qui sont généralement couplées avec des restrictions sur la durée de fonctionnement pendant la période de contrôle, le calcul de la valeur d'un élément correspondant à un équipement de ce type est difficile en raison de la forte dépendance entre les tranches de temps.

Supposons un élément correspondant au démarrage d'un équipement à t. Une manière possible de calculer sa valeur est de considérer son efficacité sur toutes les tranches temporelles de fonctionnement : $\sum_{t'=t}^{t+D(h)-1} U^a_{ht}/(D(h)P^a_M(h))$ où $D(h)$ est la durée de fonctionnement.

4.2 Solutions partiellement distribuées

Les solutions centralisées ont des limitations liées à la complexité (atténuée si elles sont basées sur des heuristiques mais le compromis est une performance plus faible). En plus, la connaissance des informations complètes sur les équipements (caractéristiques, usage prévu) risque d'être assimilé à une violation de la vie privée qui peut affecter l'acceptation de la solution par les usagers. En outre, les données doivent être envoyées à une entité centrale pour produire les signaux de contrôle qui sont ensuite envoyés aux équipements. Ceci crée une surcharge du système de communication qui peut être non négligeable et avoir un impact sur la réactivité de la solution. L'importance significative de l'entité centrale rend aussi le système vulnérable (point unique de défaillance).

Une façon naturelle de traiter les problèmes identifiés précédemment est de réduire l'importance de l'entité centrale. Dans cette section, nous nous concentrons sur les mécanismes où l'entité centrale ne contrôle que les maisons en se basant sur des informations agrégées au niveau d'une maison. Ensuite, chaque maison décide de l'allocation des équipements tout en respectant le contrôle imposé par l'entité centrale.

Dans cette section, nous allons présenter deux mécanismes de contrôle appartenant à deux familles de solutions hiérarchiques :

Solutions hiérarchiques à communication unidirectionnelle Les décisions prises par l'entité centrale (DE$_a$) sont basées sur des informations statiques qui sont recueillies à l'avance (par exemple au moment de la mise en place du contrat de service). Dans ce cas, la seule communication est liée à l'envoi des signaux de contrôle de l'entité centrale aux maisons.

Solutions hiérarchiques à communication bidirectionnelle En plus des informations statiques, l'entité centrale recueille également des informations dynamiques des maisons afin de produire des allocations plus efficaces.

La principale contribution liée à cette section est l'introduction d'une solution hiérarchique à communication bidirectionnelle basé sur une méthode de sous-gradient.

La plupart du contenu présenté dans cette section a été publié dans [KKP14, KKMP15], en collaboration avec Daniel Kofman, Fabien Mathieu et Michal Pioro.

Solution hiérarchique à communication unidirectionnelle

Comme indiqué précédemment, pour ces solutions, la DE$_a$ ne dispose que de données statiques. Sa décision ne peut donc pas être directement liée à l'amélioration de la satisfaction des usagers. Nous nous sommes concentrés sur une répartition proportionnelle de la capacité disponible sur les maisons.

Pour faire cela, les maisons sont regroupées en classes. Chaque classe reçoit une proportion de puissance constante en fonction des paramètres associés à cette classe. Pour fixer les idées, nous supposons que la répartition se fait en se basant sur la puissance maximale consommée par les équipements contrôlés dans les maisons de la classe dans des conditions de fonctionnement nominales.

Cette puissance maximale représente la puissance souscrite du contrat et est notée $L(h)$ pour chaque maison h. Les maisons d'une même classe auront la même valeur de puissance souscrite.

Supposons une capacité $C(t)$ ciblée par la DE_a pour l'instant t. A chaque tranche de temps, la DE_a attribue une puissance aux maisons proportionnelle à leur puissance souscrite. Ainsi, la limite de puissance allouée à la maison h pour la tranche de temps t est :

$$C_{ht} = \frac{L(h)}{\sum_i L(i)} C(t).$$

Sur la base de la capacité limite reçue C_{ht}, une DE_h dans une maison h décide de l'allocation des équipements en résolvant le problème centralisé (2) restreint à h, en utilisant C_{ht} au lieu de $C(t)$. Ce système de contrôle est nommé *Maximum Local d'utilité* et est notée LM(*Local Maximum*).

Solution hiérarchique à communication bidirectionnelle

Nous proposons également un système à communication bidirectionnelle qui vise un compromis entre une solution centralisée, qui fournit des performances maximales en termes d'utilité totale, et les systèmes locaux à communication unidirectionnelle, qui améliorent la scalabilité et la préservation de la vie privée. L'étude d'un tel système permet d'analyser l'importance de la communication bidirectionnelle.

Pour atteindre les objectifs de préservation de la vie privée et de scalabilité tout en limitant les informations échangées, nous proposons une décomposition primale simple du problème global GM en un problème maître, décrit par (3), et des problèmes esclaves, décrits par (4).

Problème maître

$$\max \sum_{h=1}^{H} U_h \tag{3a}$$

$$\sum_{h=1}^{H} C_{ht} = C(t), \;\; \forall\, t \tag{3b}$$

$$C_{ht} \geq 0, \;\; \forall\, h \;\; \forall\, t. \tag{3c}$$

Problèmes esclaves

Pour chaque maison h, le problème suivant est résolu :

$$U_h = \max \sum_{t=1}^{t_M} \sum_{a=1}^{A} U_{ht}^a \tag{4a}$$

$$\sum_{a=1}^{A} X_{ht}^a \leq C_{ht}, \;\; \forall\, t. \tag{4b}$$

Si les capacités C_{ht} sont connus, les problèmes esclaves (4) peuvent être résolus par les DE_h. La principale question dans ce cas est liée à la résolution du problème maître (3).

Pour résoudre ce problème, nous proposons une nouvelle heuristique appelée méthode Sous-Greedient (SG). Cette heuristique est inspirée par la méthode de sous-gradient [12], mais est adaptée pour prendre en compte les spécificités de notre modèle. En particulier, nous introduisons la notion de *Greedient* [1], inspirée par les algorithmes gloutons proposés (voir Section 1). Les Greedients seront utilisés à la place des (sous-)gradients traditionnels pour estimer la méso-pente de l'utilité pour une maison donnée.

Nous décrivons brièvement les principales étapes de SG :

- SG doit effectuer une allocation de puissance initiale.
- La DE_a transmet à chaque DE_h dans les maisons, une proposition de répartition C_{ht} $\forall t$. Chaque DE_h résout le problème (4) correspondant comme pour un système à communication unidirectionnelle. Elle renvoie l'utilité totale U_h, ainsi que le Greedient associé à la solution obtenue.

[1] Nous introduisons ce terme qui fait en même temps référence au gradient et aux algorithmes Greedy dont nous nous inspirons pour estimer la valeur d'une allocation.

- En utilisant les valeurs déclarées par les maisons, la DE_a essaie alors de proposer une meilleure répartition de la capacité.
- Le processus est itéré jusqu'à un maximum de K_{MAX} itérations. À la fin, la meilleure répartition trouvée est utilisée.

Nous donnons maintenant les détails supplémentaires nécessaires pour avoir une vue complète de la solution.

Allocation initiale

La première allocation (avant le premier retour des maisons) est nécessairement une répartition basée sur des informations statiques. Nous proposons d'utiliser une répartition round-robin de la capacité en allouant la capacité souscrite $L(h)$ au maximum de maisons et en coupant le reste. Le processus est itéré sur la période de contrôle en considérant une sélection basée sur une liste cyclique des maisons. L'intérêt pour SG d'une telle allocation initiale (par exemple par rapport à LM) est de casser les symétries possibles entre les maisons et de donner une diversité initiale qui aidera à trouver de bons Greedients.

Greedient

Nous définissons le greedient g_{ht} comme le meilleur rapport possible entre l'amélioration de l'utilité de la maison h et l'amélioration des capacités à une tranche de temps t. Formellement, si $U_h'(\Delta C_t)$ représente la meilleure utilité possible pour la maison h quand la capacité est augmentée de ΔC_t à t, nous aurons :

$$g_{ht} := \max_{\Delta C_t > 0} \frac{U_h'(\Delta C_t) - U_h}{\Delta C_t}.$$

Pour calculer le greedient d'une maison, nous définissons le greedient g_{ht}^a d'un équipement a comme suit : pour une C_{ht} donnée, $C_{ht}^0 \geq 0$ représente la capacité inutilisée par la maison h à t pour l'allocation optimale. $U_h^{a\prime}(\Delta C_t)$ représente l'utilité maximale pour l'équipement a si une capacité supplémentaire allant jusqu'à ΔC_t est ajoutée à sa puissance allouée. Ainsi, nous aurons :

$$g_{ht}^a := \max_{\Delta C_t > 0} \frac{U_h^{a\prime}(C_{ht}^0 + \Delta C_t) - U_h^a}{\Delta C_t}.$$

Il est facile de voir que le greedient d'une maison est le greedient de son équipement le plus efficace : $g_{ht} = \max_a g_{ht}^a$.

Amélioration de la solution

Pour mettre à jour la répartition de la puissance à la k-ième itération, la DE_a effectue les opérations suivantes :

- Elle calcule d'abord les valeurs $\alpha_k g_{ht}$ $\forall h \forall t$. Ces valeurs représentent l'augmentation potentielle de C_{ht}. Le paramètre α_k est appelé *taille du pas* et sera décrit plus loin.
- Elle ajuste ensuite les nouvelles valeurs de C_{ht} afin qu'elles restent positives et que leur somme soit égale à la capacité $C(t)$.

Pour la phase d'ajustement, il est important de traiter les cas où $\alpha_k g_{ht}$ est plus grand que la capacité disponible $C(t)$ ou même la puissance souscrite $L(h)$ de la maison h. Nous avons donc limité $\alpha_k g_{ht}$ au minimum entre la limite de puissance de la plus petite maison ($L_m := \min_h L(h)$) et la capacité du système $C(t)$. Nous définissons donc $\beta_{kht} = \min(\alpha_k g_{ht}, Lm, C(t))$.

Ensuite, pour chaque t, une valeur positive commune λ_t est soustraite de C_{ht} pour maintenir la somme des allocations égales à la capacité totale $C(t)$. Pour éviter que les maisons à faible C_{ht} soient trop impactées et des allocations négatives due à l'opération de soustraction, un sous-ensemble I_t de maisons sera «protégé» afin que leurs allocations ne diminuent pas. Ceci est fait comme suit, en commençant par $I_t = \emptyset$:

- Nous calculons λ_t de telle sorte que les valeurs

$$C'_{ht} = \begin{cases} C_{ht} + \max\{\beta_{kht} - \lambda_t, 0\} \text{ if } h \in I_t, \\ C_{ht} + \beta_{kht} - \lambda_t \text{ sinon,} \end{cases} \tag{5}$$

  somment à $C(t)$. Voir [79, 40] pour plus de détails.

- Nous protégeons (ajoutons à I_t) toutes les maisons qui obtiennent une allocation C'_{ht} négative.

- Nous parcourons les deux étapes ci-dessus jusqu'à ce que tous les C'_{ht} de l'équation (5) soient positifs. La DE$_a$ propose alors C'_{ht} comme une nouvelle répartition envisageable.

Remarques : Bien que la solution décrite ici soit valable pour une hiérarchie à 2 niveaux (DE$_a$, DE$_h$), elle peut être généralisée à m niveaux. Le greedient pour une entité à un niveau donné sera le greedient maximal de ses enfants. La phase d'ajustement peut aussi prendre en compte des contraintes supplémentaires de capacité à chaque point d'agrégation en les supposant comme des limites statiques de puissance.

La solution proposée peut fonctionner de manière asynchrone dans le sens où elle ne nécessite pas que toutes les maisons communiquent simultanément. En fait, dès qu'au moins deux maisons répondent, une redistribution locale peut être faite sans avoir à attendre une réponse des autres maisons : nous avons juste besoin de limiter le problème au sous-ensemble correspondant de maisons, en utilisant leur allocation cumulée comme capacité totale disponible.

Taille du pas

La taille du pas α_k pour chaque itération k est un paramètre crucial de la solution. En effet, le choix de tailles de pas appropriées est la clé pour accélérer la résolution. Intuitivement, une grande valeur de α_k permettra de mettre en marche des équipements consommateurs et une valeur plus basse est plus adaptée aux équipements à faible consommation.

Parmi les méthodes possibles pour définir la taille du pas à chaque itération, nous avons étudié dans le mémoire (Chapitre 5) certaines méthodes adaptées au problème qu'on cherche à résoudre. Nous allons présenter ici une des méthodes étudiées inspirée de la règle du pas à longueur constante (see [12]). La méthode consiste à prendre une taille de pas de la forme $\alpha_k = \frac{a}{||g_{ht}||_2}$, où $||g_{ht}||_2$ est la norme euclidienne du vecteur de greedients. Cette méthode assure une longueur de pas constante ($||C'_{ht} - C_{ht}||_2 = a$) en l'absence de contraintes de capacité.

La valeur du paramètre a est critique pour la performance du système. Actuellement, elle est réglée manuellement de manière à fournir le meilleur résultat, mais nous croyons que l'estimation automatique de la meilleure valeur est une piste prometteuse pour les travaux futurs.

4.3 Solutions distribuées

Dans la section précédente, nous avons présenté deux familles de systèmes de contrôle hiérarchique : une à communication unidirectionnelle, et une à communication bidirectionnelle. La communication bidirectionnelle permet d'adapter les allocations en prenant en compte les besoins des maisons, ce qui doit améliorer la performance de l'allocation. Cependant, cette technique ne résout pas le problème du point unique de défaillance comme l'entité centrale joue un rôle important dans la mise à jour des allocations. Dans cette section, nous visons à proposer des solutions qui auront de meilleures performances pour toutes les caractéristiques souhaitées : la scalabilité, l'efficacité et la préservation de la vie privée. Ceci est réalisé en permettant la communication entre les maisons. En effet, en cas d'anonymat, les maisons peuvent coordonner les uns avec les autres afin d'améliorer leur qualité d'expérience sans compromettre leur vie privée. Cela donne une solution qui est résiliente au point de défaillance unique.

Formulation du problème

Nous visons à proposer des techniques de contrôle qui permettent aux maisons d'échanger des blocs de puissance afin d'améliorer l'efficacité de leurs allocations. La base de notre système est simple : un ensemble de maisons tente d'améliorer leur qualité d'expérience en résolvant un problème d'optimisation local. Cette opération centrale est appelée une *activation*. En cas de succès, une activation se traduira par un échange multilatéral de blocs de puissance correspondant à la solution du problème. Pour cela, nous supposons que les maisons utilisent le service à travers une plateforme d'échange qui garantie l'anonymat. Lorsqu'une maison utilise la plateforme, elles peut communiquer avec un ou plusieurs pairs afin d'explorer les possibilités d'échange d'énergie.

Appelons N le nombre de maisons impliquées dans la coordination. Une (ou plusieurs) des maisons participant à la coordination va résoudre le problème de maximisation du bien-être social tout en tenant compte des informations fournies par les autres $N-1$ maisons. L'allocation sera déterminée de telle sorte que les maisons chercheront à se conformer à la somme de leurs limites de capacité. Le problème d'optimisation est formulé par les équations (6).

$$\max_{X^a_{h_i t}, x^a_{h_i t}} \sum_{t=1}^{t_M} \sum_{i=1}^{N} \sum_{a=1}^{A} U^a_{h_i t} \tag{6a}$$

s.t.

$$\sum_{i=1}^{N} \sum_{a=1}^{A} X^a_{h_i t} \leq \sum_{i=1}^{N} C(h_i, t) + c_r(t), \quad \forall\, t \tag{6b}$$

$$P^a_m(h_i) x^a_{h_i t} \leq X^a_{h_i t} \leq P^a_M(h_i) x^a_{h_i t}, \quad \forall\, t, \forall\, i, \forall\, a \tag{6c}$$

$$x^a_{h_i t} \in \{0, 1\}, \quad \forall\, t, \forall\, i, \forall\, a. \tag{6d}$$

Dans cette formulation, t_M désigne la période d'optimisation. L'Équation (6b) impose une consommation commune qui ne dépasse pas la capacité disponible. Cette capacité se compose de deux éléments :

- Le paramètre $C(h_i, t)$ représente la capacité assignée à la maison h_i au temps t
- Le paramètre $c_r(t)$ représente la capacité disponible supplémentaire que les maisons peuvent utiliser (plus de détails ci-dessous).

La sincérité des usagers peut être garantie en exigeant que plusieurs des maisons concernées résolvent le problème d'optimisation et comparent la cohérence des solutions obtenues.

Une fois (6) résolu, la solution est proposée à tous les usagers participant à l'échange. Si tout le monde accepte la solution, l'activation est réussie et la solution devient la nouvelle répartition effective. Dans le cas contraire, l'activation échoue et l'allocation des maisons reste inchangée.

La solution complète consiste simplement à exécuter des activations plusieurs fois sur plusieurs ensembles de maisons de façon asynchrone jusqu'à ce qu'une solution acceptable soit atteinte. Cela dit, nous avons encore besoin de définir quelques détails du protocole pour avoir une description complète de la solution :

Contraintes locales Les utilisateurs sont-ils d'accord pour dégrader leur propre qualité d'expérience afin d'accroître la qualité d'expérience globale ?

Ordonnancement Comment l'ensemble des pairs impliqués dans une activation est-il décidé ?

Gestion de l'excès Pour une allocation donnée, il est probable que certaines capacités restent inutilisées. Comment pouvons-nous les utiliser (à savoir, le terme $c_r(t)$ dans (6b)) ?

Nous n'allons pas élaborer sur les méthodes de gestion de l'excès : nos analyses numériques n'ont pas permis de montrer une différence significative de performance des différentes méthodes.

Contraintes locales

Le problème présenté précédemment n'exprime pas toutes les contraintes locales liées aux usagers pour l'approbation de l'échange. En pratique, un usager peut ne pas être prêt à avoir sa propre qualité d'expérience abaissée. Ainsi, des contraintes supplémentaires au problème d'optimisation doivent être imposées.

Nous nous intéresserons à des contraintes locales qui exprimeront trois types de comportement des usagers, à savoir égoïstes, coopératifs et intermédiaires. Ces comportements sont décrits comme suit :

- *Les usagers égoïstes* n'accepteront aucun échange qui n'améliore pas leur utilité perçue ;
- *Les usagers coopératifs* acceptent le résultat de l'optimisation si la solution améliore l'utilité totale perçue du groupe d'usagers concerné ;
- *Les usagers intermédiaires* coopèrent au niveau de leur utilité de confort, mais sont égoïstes par rapport aux besoins vitaux.

Nous formalisons maintenant les contraintes supplémentaires correspondant aux trois cas de comportement supposés. Le changement d'utilité totale pour la maison h_i si l'activation est réussie est notée ΔU_{h_i}.

Les contraintes locales supplémentaires sont exprimées en ΔU_{h_i} et sont formulées comme suit :

- Un usager égoïste h_i imposera la contrainte $\Delta U_{h_i} \geq 0$.
- Un usager coopératif h_i n'imposera aucune contrainte sauf le fait que la solution soit strictement meilleure que celle active ($\sum_{i=1}^{N} \Delta U_{h_i} \geq 0$).
- Un usager intermédiaire h_i imposera la contrainte $\Delta U_{vh_i} \geq 0$.

Ces contraintes sont ajoutées à la formulation du problème défini par l'équation (6) selon le type de comportement considéré pour chaque usager. Un échange est approuvé lorsque l'utilité est globalement améliorée pour les usagers. Dans le présent travail, nous supposons que toutes les maisons ont le même comportement [2]. Nous appelons donc Selfish Distributed Maximum (SDM), Cooperative Distributed Maximum (CDM) et Intermediate Distributed Maximum (IDM), les solutions qui découleront de la coordination entre maisons dans le cas des usagers égoïstes, coopératifs et intermédiaires, respectivement.

[2] Le framework permet des comportements hétérogènes par exemple selon que l'utilisateur ait signé pour un service premium ou non.

Ordonnancement

Dans notre solution proposée, chaque maison contacte régulièrement d'autres maisons. Pour des fins d'évaluation, nous approchons ce comportement par un ordonnanceur round-robin, un *round* correspondant à une séquence où chaque maison a proposé une activation.

Le choix des maisons qui participent à une activation est crucial. En particulier, la taille du problème est important, pour des raisons de complexité et de sécurité. En effet, on suppose que, lorsque les maisons coordonnent, elles ont besoin de partager leurs préférences avec leurs pairs par le biais de la plateforme d'échange. En garantissant l'anonymat, cet échange d'informations peut se faire sans divulguer l'identité des maisons. Toutefois, si le nombre de maison augmente par rapport au nombre de maisons utilisant la plateforme, la probabilité d'apprentissage sur les maisons augmente. En outre, la complexité de calcul augmente avec le nombre de la maison. En plus, plus le nombre est bas, plus la plateforme est flexible comme les communications seront restreintes.

Pour cette raison, nous proposons de limiter la taille à $N = 3$ maisons, à savoir quand une maison déclenche une activation, elle contacte 2 autres maisons. Pour équilibrer l'exhaustivité et le caractère aléatoire des groupes considérés, nous proposons qu'une maison est contactée de manière round-robin déterministe et une au hasard (parmi toutes les maisons participantes à la plateforme). Nous considérons alors deux ordonnanceurs simples selon que l'activation proposée repose sur les trois maisons ou seulement une partie d'entre eux.

Ces techniques sont nommés activation directe et meilleur activation.

Activation directe Pour un premier ordonnanceur simple, nous supposons que toutes les maisons contactées participent à l'activation. Si toutes les maisons sont satisfaites du résultat de l'optimisation jointe, une modification des allocations de puissance est directement mise en place. Cet ordonnanceur est donc appelé «Direct Activation» (Activation directe en français) et noté *DA*.

Meilleure activation Au lieu d'activer directement l'échange d'énergie avec toutes les maisons contactées, l'activation ne sera effectuée que sur un sous-ensemble. Pour $N = 3$, le comportement suivant est obtenu : l'initiateur de l'activation va calculer le résultat de l'activation avec le premier pair contacté et le résultat de l'activation avec le second, et de choisir la meilleure option (en terme d'utilité globale) entre ces deux options. Cette approche est inspirée par le principe de *power of two choices* [4] et vise à réduire au minimum la taille des problèmes à résoudre. Cette technique est appelée Meilleur Activation. Elle est notée *BA*.

5 Analyses numériques

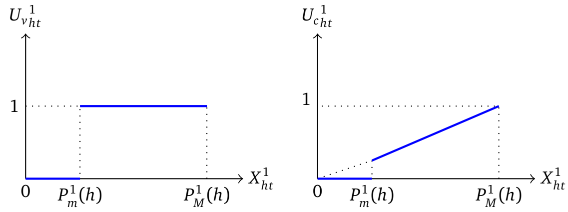
Dans cette section, nous allons évaluer la performance des solutions proposées en supposant un cas d'usage particulier. En effet, plusieurs cas d'usage sont présentés dans la thèse. Cependant, nous allons nous contenter de présenter un seul cas représentatif mettant en avance les conclusions tirées.

Les paramètres du système sont choisis de manière à avoir un cas neutre : aucune préférence de priorité n'est exprimée entre les maisons et les équipements. Donc, nous supposons un cas où toutes les maisons ont la même utilité maximale (à savoir, pas de différenciation de service). En outre, l'utilité maximale de la maison est répartie de manière équitable sur les équipements (à savoir, tous les équipements ont la même utilité totale maximale).

5.1 Paramètres

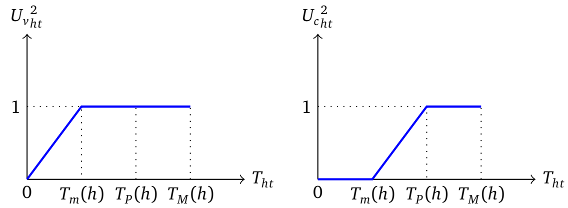
Pour étudier la performance des systèmes de contrôle pour plusieurs valeurs de capacité, nous considérerons les paramètres suivants :

- Une durée de tranches temporelles de 5 minutes.



(a) Utilité vitale (b) Utilité de confort

FIGURE 1 – Utilité de la puissance d'éclairage



(a) Utilité vitale (b) Utilité de confort

FIGURE 2 – Utilité liée à la température

- Deux types d'équipements (A = 2) : un système d'éclairage (indice $a = 1$) et du chauffage (indice $a = 2$). Les fonctions d'utilité des deux équipements ont une composante vitale et une composante de confort.
- L'utilité vitale de l'éclairage est entièrement satisfaite quand la puissance minimale $P_m^1(h)$ est attribuée, alors que l'utilité de confort croit linéairement entre $P_m^1(h)$ et $P_M^1(h)$ (voir Figure 1).
- Pour le chauffage, l'utilité vitale augmente linéairement jusqu'à l'atteinte d'une température minimale tolérable $T_m(h) := 15°C$, tandis que l'utilité de confort croit linéairement entre $T_m(h)$ et $T_P(h) := 22°C$ (voir Figure 2). Pour garantir le bien être des usagers, la température ne doit pas dépasser $T_M(h) := 30°C$.
- Nous supposons une température extérieure constante $T_e(t) = 10°C \ \forall t$ et une température initiale $T_0(h) = 22°C \ \forall h$.
- L'évolution de la température dans les maisons suit un modèle simplifié de conductivité/capacité dont la dynamique est la suivante :
  $T_{ht} = T_{h(t-1)} + F(h)X_{ht}^2 + G(h)(T_e(t) - T_{h(t-1)})$.
- Le système est constitué de $H = 100$ maisons.

Nous supposons que la puissance totale disponible est constante sur la période d'optimisation, $C(t) = C$. Nous analysons le modèle pour différentes valeurs de C, allant de faible à pleine capacité (tous les équipements peuvent être utilisés). Bien que ce modèle soit assez simple (deux types d'appareils, des valeurs constantes pour C et T_e), nous pensons qu'il est suffisant pour donner un aperçu des exigences de chaque système de contrôle proposé.

L'analyse numérique des différents systèmes de contrôles (GM, LM, SG et des solutions distribuées) a été effectuée en résolvant les problèmes par IBM ILOG CPLEX v11.2 ([45]). Étant donné que certains problèmes prennent beaucoup de temps pour une résolution locale, nous avons fixé une conditions d'arrêt supplémentaires qui limite le temps d'une résolution à 7 heures. L'heuristique gloutonne $RTGreedy$ est implémentée en C++. Tous les programmes sont exécutés sur un serveur Linux CentOS avec 2x4 cœurs CPU, 16 Go de RAM et 2,66 Ghz de vitesse d'horloge.

5.2 Scénarios de maisons

Nous analysons les systèmes de contrôle proposés en supposant des scénarios de maisons homogènes et hétérogènes. Les paramètres pour chaque scénario sont les suivants.

Maisons homogènes

Pour ce scénario, toutes les maisons ont les mêmes caractéristiques d'équipements et appartiennent à la même classe désignée par classe 1. Comme indiqué dans les tableaux 1, 2 et 3, nous définissons la classe 1 par les puissances (minimale et maximale) pour chaque equipement (en Watts) : système d'éclairage [50, 1000], système de chauffage [1000, 4000]. Les paramètres thermiques $F(h)$ et $G(h)$ sont pris égal à 0,0017 et 0,075 respectivement.

Maisons hétérogènes

Pour ce scénario, les maisons appartiennent à deux classes. Nous considérons que 50 maisons appartiennent à la classe 1 présentée précédemment. Les 50 autres maisons sont de classe 2 pour laquelle les équipements ont les puissances (minimale et maximale) suivantes (en Watts) : système d'éclairage [50, 500], système de chauffage [1000, 2000]. En ce qui concerne les paramètres thermiques $F(h)$ et $G(h)$, on suppose que leurs valeurs sont égales à 0,0008 et 0,0365 respectivement. La valeur de $G(h)$ signifie que sans chauffage, la température intérieure diminue de $0,365°C$ sur une tranche de temps pour la classe 2 comparée à une $0.75°C$ pour la classe 1, quand la différence de température avec la température extérieure est égale à $10°C$. Cela signifie que les maisons de la classe 2 sont mieux isolées que celles de la classe 1. Les paramètres définissant la classe 2 sont résumés dans les tableaux 1, 2 et 3.

Type	$P_m^1(h)$	$P_M^1(h)$
1	50	1000
2	50	500

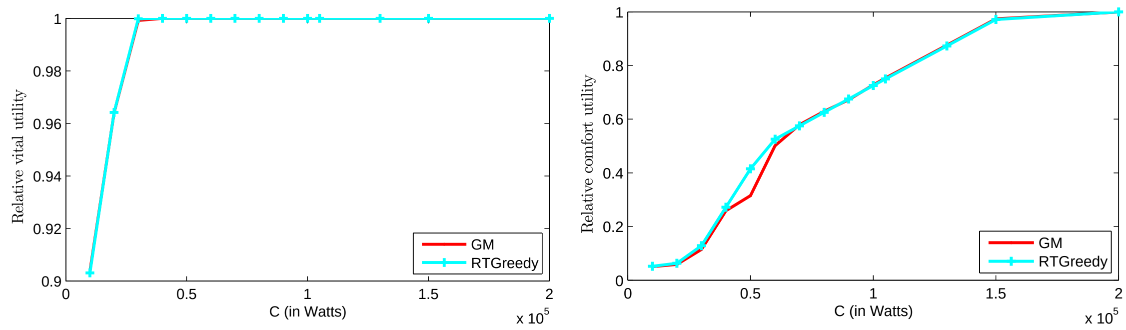
TABLE 1 – Paramètres de l'éclairage

Type	$P_m^2(h)$	$P_M^2(h)$	$F(h)$	$G(h)$
1	1000	4000	1.7×10^{-3}	7.5×10^{-2}
2	1000	2000	0.8×10^{-3}	3.65×10^{-2}

TABLE 2 – Paramètres du chauffage

Classe	Type d'éclairage	Type de chauffage	Puissance maximale $L(h)$
1	1	1	5000
2	2	2	2500

TABLE 3 – Paramètres des maisons



(a) Utilité vitale

(b) Utilité de confort

FIGURE 3 – Utilité relative en fonction de la capacité disponible pour des solutions centralisées (cas homogène)

5.3 Solutions centralisées

L'évaluation des algorithmes centralisés est effectuée en comparant les performances à ceux de GM. Comme il est basé sur le contrôle direct et les données les plus fines, GM fournit l'utilité totale maximale sur l'horizon d'optimisation (performance maximale). Nous allons donc le considérer comme la référence pour évaluer la performance de l'algorithme glouton centralisé proposé ainsi que les autres algorithmes des autres familles de solutions. Dans ce cas, nous considérons une période d'optimisation fixée à $t_M = 100$ tranches de temps ($\approx$ 8 heures).

Résultats du scénario homogène

La Figure 3 montre les résultats pour le cas homogène.Pour faciliter l'analyse, nous comparons l'utilité totale obtenue sur toute la période d'optimisation à l'utilité maximale réalisable. Par conséquent, nous normalisons l'utilité totale en divisant par le nombre de maisons, le nombre de tranches de temps et le nombre de classe d'équipements. La performance est donc exprimée par l'utilité totale normalisée que nous appelons l'utilité relative. Seuls les points dans les graphiques ont été calculés.

Pour comparer les performances, nous rappelons l'ordre lexicographique entre l'utilité vitale et celle de confort. En effet, lorsque deux points (correspondant à deux mécanismes de contrôle) pour une certaine valeur de la capacité C sont comparés, nous comparons d'abord les utilités vitales avant de comparer celles de confort.

En regardant la Figure 3, nous pouvons constater une performance similaire de GM et de $RTGreedy$. En fait, nous pouvons même remarquer que GM ne produit pas l'utilité maximale pour la capacité $C = 0.5 \times 10^4$. Ceci est dû à la complexité d'une résolution exacte à l'aide de CPLEX qui est arrêtée si une solution optimale n'est pas fournie dans un temps raisonnable. Cela montre l'intérêt des heuristiques simples capables d'atteindre de bonnes performances tout en réduisant le temps de résolution.

Résultats du scénario hétérogène

Une analyse du cas hétérogène montre également des performances similaires entre GM et $RTGreedy$ même si les solutions obtenues peuvent être différentes. Nous avons observé une difference notable dans les stratégies d'allocation pour des capacité faibles du système où GM favorise les maisons efficaces alors que $RTGreedy$ favorise les maisons moins isolées à cause d'une surestimation de l'importance du chauffage.

5.4 Solutions partiellement distribuées

Nous comparons maintenant les performances des solutions hiérarchiques à la performance d'une solution globale. Pour le problème sous-Greedient, nous fixons le nombre maximum d'itérations à $K_{MAX} = 100$ et nous prenons $a = 6000$. La solution est désignée par $SG-2$. Nous analysons les performances en considérant un scénario homogène et un autre hétérogène. Dans ce cas, nous considérons une période d'optimisation fixée à $t_M = 60$ tranches de temps (5 heures).

Résultats du scénario homogène

Nous analysons le comportement des algorithmes partiellement distribués en regardant différentes capacités globales du système. Ces capacités sont choisies de telle sorte que l'allocation initiale LM ne puissent pas fournir une allocation optimale globale. Comme pour les solutions centralisées, nous utiliserons dans nos analyses l'utilité relative qui atteint sa valeur maximale de 1 lorsque l'utilité totale maximale est atteinte (pour vital ou pour confort).

La Figure 4 présente l'utilité relative perçue par les maisons pour plusieurs valeurs de capacité globale du système.

Une valeur d'intérêt pour l'utilité vitale est 0,875 qui correspond à des situations où toutes les maisons ont l'éclairage vital ($P_m^1 = 50$ W) mais aucun n'a la puissance nécessaire pour le chauffage ($P_m^2 = 1000$ W).

La solution optimale GM, réalise l'utilité vitale maximale même pour de très faibles capacités (jusqu'à 4×10^4), grâce à sa capacité de trouver une allocation qui permet à toutes les maisons d'utiliser le chauffage pour une durée suffisante pour remplir les besoins vitaux. Basés sur les résultats de GM, nous pouvons mesurer l'écart entre l'allocation optimale et les allocations obtenues par les solutions partiellement distribuées .

Une allocation statique, par LM ne permet que tardivement d'augmenter l'utilité vitale au-dessus du seuil de 0,875. Le chauffage ne peut être utilisé qu'à partir de $C = 10^5$ (1000 W par maison). L'utilité vitale maximale ne peut être atteinte que pour des capacités supérieures ou égales à $C = 105 \times 10^3$ (1050 W par maison). Toutefois, LM donne de bonne performance pour les valeurs de capacité disponible suffisamment élevées.

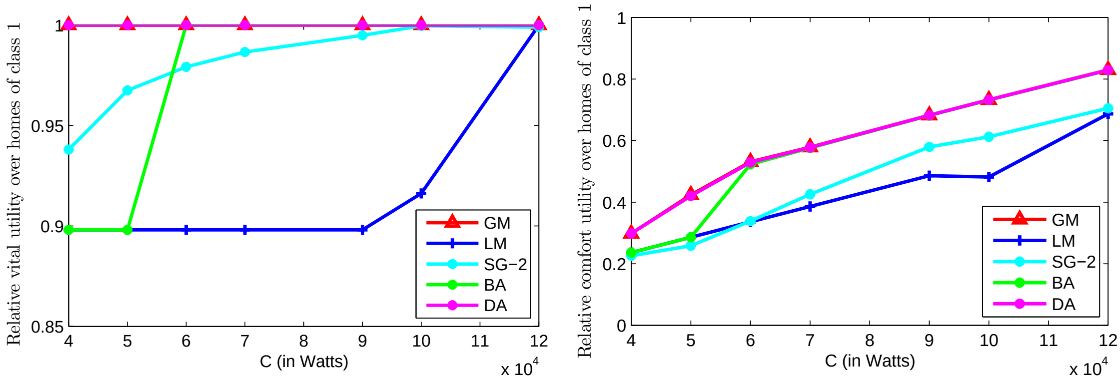
Comme prévu, notre proposition, $SG-2$, a des performance intermédiaires par rapport aux deux algorithmes GM et LM. Il est capable d'améliorer l'utilité vitale de maisons pour des valeurs inférieures à $C = 10^5$, même s'il ne parvient pas à atteindre les performances de GM. En ce qui concerne l'utilité de confort, il est à égalité avec LM, même dans les situations où il fournit le chauffage vital tandis que LM ne le fait pas.

Résultats du scénario hétérogène

Pour le scénario hétérogène, les résultats sont illustrés dans la Figure 5. Nous pouvons observer le même comportement que celui du scénario homogène où $SG-2$ offre une performance meilleure que LM, mais qui peut ne pas atteindre celle de GM pour toutes les valeurs de capacité.

En fait, pour l'utilité vitale, les résultats de GM sont à peu près similaires pour les deux classes au cas homogène, avec une valeur maximale obtenue même pour de faibles capacités (jusqu'à 4×10^4). Pour l'utilité de confort, cependant, on remarque que GM conduit à de meilleures valeurs pour la classe 2 par rapport à la classe 1. Cela est dû au fait que les maisons de la classe 2 sont d'une performance énergétique meilleure, donc une fois l'utilité vitale assurée pour tous, il est plus efficace d'allouer l'énergie aux maisons de la classe 2.

La même raison explique la mauvaise performance de LM. Rappelons que l'allocation statique est proportionnelle à la puissance maximale $L(h)$ des maisons. Donc, pour une capacité donnée, les maisons de la classe 1 vont obtenir plus de puissance que celles de la classe 2. En conséquence, même si la performance de la classe 1 est satisfaisante, la performance de la classe 2 est terrible malgré la meilleure performance énergétique des maisons de la classe 2. En particulier, la capacité



(a) Utilité vitale

(b) Utilité de confort

FIGURE 4 – Utilité relative en fonction de la capacité disponible (cas homogène)

requise par les maisons de la classe 2 pour atteindre l'utilité vitale maximale est très élevée supérieure à $C = 1,2 \times 10^5$, ce qui correspond à 1200 W par maison (quelle que soit la classe).

Nous constatons que comparée au cas homogène, la performance de notre solution $SG-2$ est maintenant plus proche de GM que de LM. En particulier, $SG-2$ parvient à profiter de l'hétérogénéité pour atteindre des valeurs d'utilité vitales élevées plus rapidement comparé au cas homogène (Figure 3). En ce qui concerne l'utilité de confort, elle reste en dessous des valeurs obtenues par GM mais parvient à donner des valeurs descentes pour les deux classes, ce qui donne un net avantage sur LM (en particulier concernant les maisons de la classe 2).

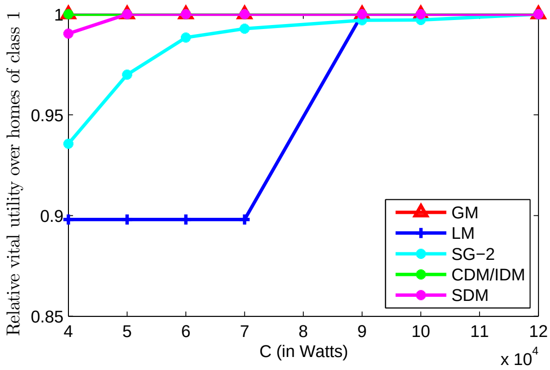
5.5 Solutions distribuées

Dans cette analyse, nous supposons une période d'optimisation fixée à $t_M = 60$ tranches de temps (5 heures). Pour les algorithmes distribués, les maisons vont tenter des activations avec d'autres maisons de manière asynchrone (voir Section 1). Pour des fins d'analyse, nous définissons :

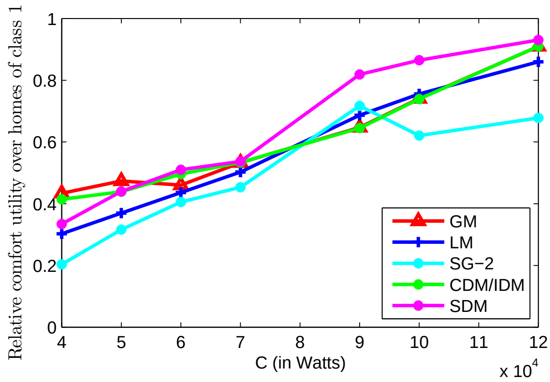
- une *allocation de puissance initiale* : N'importe quel algorithme qui peut déterminer des capacités initiales de telle sorte que la somme de ces capacités soit égale à la capacité globale disponible sur la période d'optimisation, peut être considéré comme solution initiale. Dans cette analyse numérique, nous choisissons une allocation initiale qui suit LM pour montrer comment les systèmes de contrôle distribués peuvent modifier une allocation simple.

- Une *condition d'arrêt* pour le problème d'optimisation. Même si les algorithmes proposés peuvent fonctionner en continu en tâche de fond, les paramètres considérés ici sont statiques. Nous avons donc besoin d'une condition d'arrêt : si l'amélioration de l'utilité relative entre deux rounds consécutifs est inférieure à 10^{-2} ou au bout de 100 rounds, l'algorithme s'arrête. Ce choix est motivé par le désir d'explorer une amélioration possible, même si elle est faible, tout en limitant cette exploration.

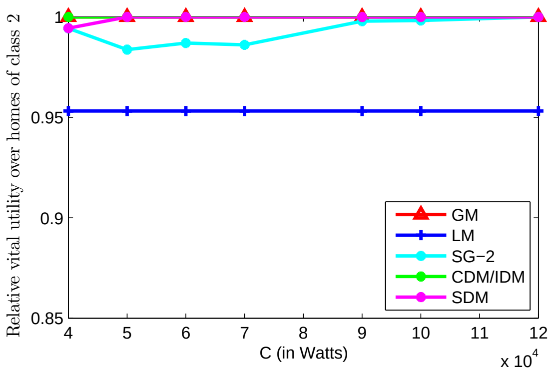
Nous présentons maintenant les résultats des algorithmes distribués pour toutes les contraintes locales et ordonnanceurs déjà présentées. Les algorithmes sont différenciés dans les légendes des graphiques sur la base des paramètres pertinents : les contraintes locales (à savoir, CDM, SDM ou IDM), ordonnanceur (à savoir, *DA* ou *BA*).



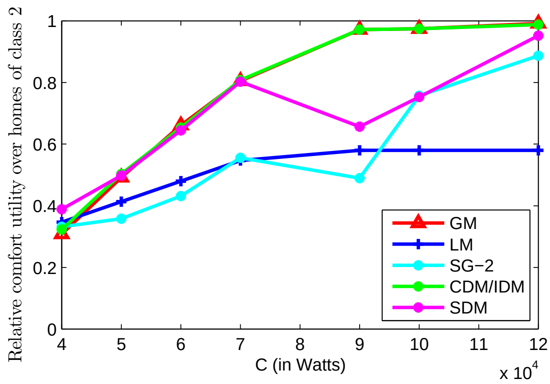
(a) Utilité vitale pour les maisons de classe 1



(b) Utilité de confort pour les maisons de classe 1

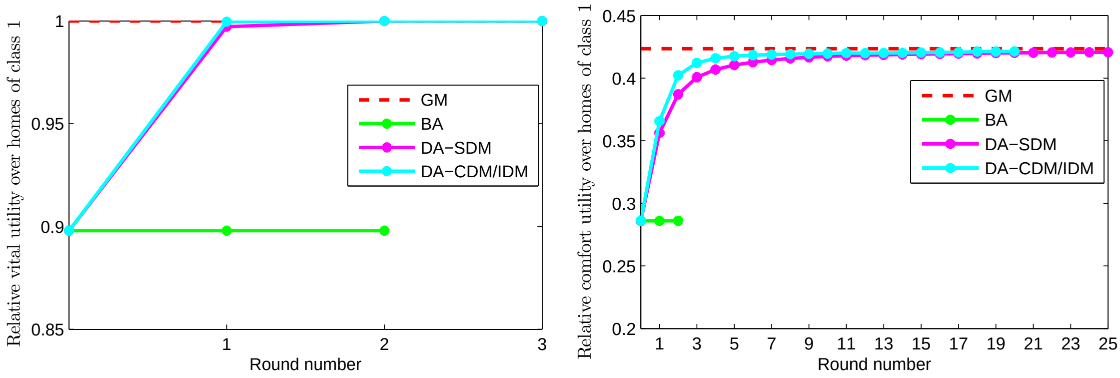


(c) Utilité vitale pour les maisons de classe 2



(d) Utilité de confort pour les maisons de classe 2

FIGURE 5 – Utilité relative en fonction de la capacité disponible (cas hétérogène)



(a) Utilité vitale

(b) Utilité de confort

FIGURE 6 – Utilité relative en fonction du nombre de rounds pour une capacité de 50000 Watts (cas homogène)

Remarque : Dans ce qui suit, nous utilisons une gestion distribuée de l'excès : (la délégation des tranches de temps est affectée aux maisons de manière aléatoire ; la maison responsable d'une tranche de temps doit mettre à jour l'excès relatif à cette tranche après chaque activation réussie.

Résultats du scénario homogène

La Figure 4 représente aussi les résultats des algorithmes distribués. Nous pouvons voir que les performances des algorithmes distribués atteignent celles de *GM* pour toutes les capacités testées, à l'exception du cas où l'ordonnanceur *BA* est considéré sur des faibles capacités ($C = 5 \times 10^4$W ou moins). En effet, pour une capacité globale de $C = 5 \times 10^4$W, qui correspond à 500W par maison, l'optimisation sur deux maisons en utilisant l'ordonnanceur *BA* ne peut garantir le maximum vital. En effet, la puissance totale des deux maisons est équivalente à 1000W et est entièrement allouée à l'éclairage. Donc, même si la capacité totale correspond exactement à ce qu'un chauffage consomme pour fonctionner, ils ne seront pas exploités car les besoins vitaux pour l'éclairage seront touchés (l'éclairage est considéré plus rentable que le chauffage). En travaux futurs, il serait intéressant d'étudier la possibilité de considérer toute capacité utilisée pour l'amélioration du confort comme un nouveau type de capacité «résiduelle» afin de répondre aux besoins vitaux.

La Figure 6 présente l'évolution de l'utilité vitale totale en fonction du nombre de rounds pour une capacité $C = 5 \times 10^4$W. Nous pouvons voir que les systèmes distribués utilisant *DA* (3 maisons impliquées) sont capables d'atteindre la performance de *GM* en moins de 20 itérations. Cependant, nous pouvons remarquer que l'égoïsme joue un rôle dans le ralentissement de la convergence des algorithmes. Nous pouvons voir aussi que les algorithmes utilisant *BA* s'arrêtent après le second round car aucune amélioration n'est perçue (voir la condition d'arrêt ci-dessus).

En ce qui concerne le temps d'exécution, la Figure 7 montre la durée de résolution relative par round pour les algorithmes testés. Les valeurs indiquées correspondent à une normalisation des temps d'exécution par round par rapport au temps pris par *GM*. Ainsi, lorsque le temps relatif est égal à 1, le temps d'exécution d'un round est égal au temps d'exécution de *GM*. Nous pouvons voir que *BA* a généralement une résolution plus rapide que celle de *DA*. Cela est principalement lié au nombre de maisons impliquées dans l'optimisation (2 contre 3 maisons). En outre, des usagers égoïstes affectent le temps d'exécution d'un round pour un ordonnanceur donné. Il peut sembler que *GM* a une bonne durée de résolution. Cependant, ces temps sont très dépendants de la taille du problème, ainsi que des paramètres du système qui affectent la complexité de l'algorithme. En outre, l'intérêt des algorithmes distribués et leur vitesse de résolution deviennent plus significatifs lorsque le nombre total de maisons augmente. Par ailleurs, en pratique, le temps de résolution des algorithmes distribués est en fait plus faible que le temps montré sur la figure. En effet, dans un contexte réaliste, au cours d'un round, les maisons peuvent optimiser leurs allocations en parallèle.

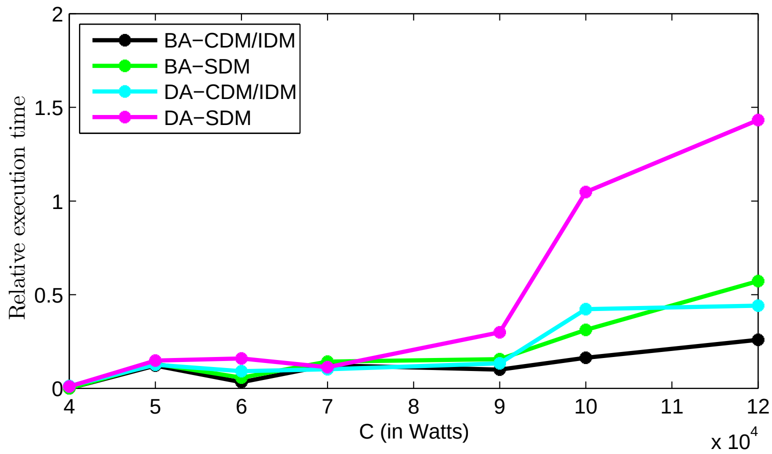


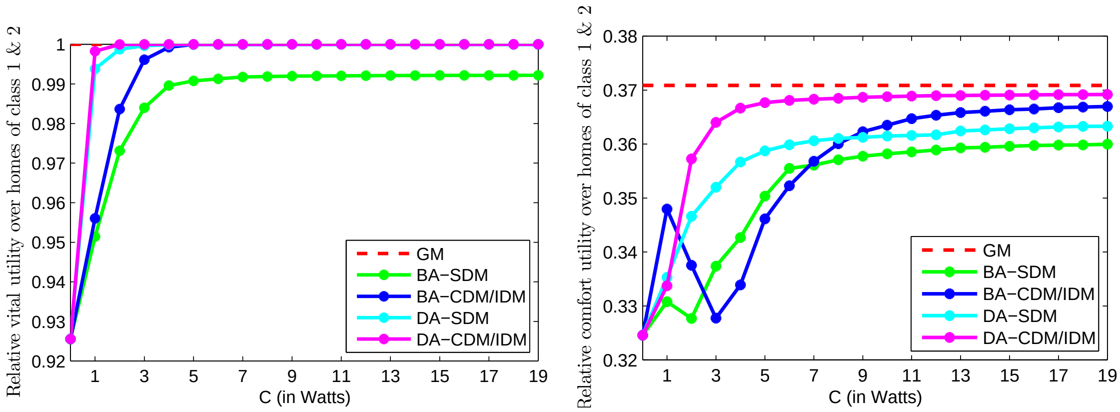
FIGURE 7 – Temps de calcul relatif (cas homogène)

Résultats du scénario hétérogène

La Figure 5 montre les résultats pour le scénario hétérogène. Tous les algorithmes distribués qui reposent sur une comportement coopératif ou intermédiaire atteignent des performances proche de l'optimal pour toutes les capacités testées. Un comportement égoïste est sous-optimal pour les faibles valeurs de capacité (sous-optimalité vitale) et des valeurs élevées (sous-optimalité de confort). Par exemple, si nous regardons les capacités élevées ($C \geq 9 \times 10^4$), nous pouvons remarquer qu'un comportement égoïste a tendance à favoriser le confort des usagers de la classe 1 en raison de l'allocation initiale considérée. En effet, après avoir rempli leurs besoins vitaux, les maisons de la classe 1 égoïstes ne sacrifieront pas leur confort afin d'améliorer l'utilité des maisons de la classe 2. Il en résulte une solution sous-optimale.

Analysons maintenant plus en détail ce qui se passe pour une capacité basse. La Figure 8 illustre la vitesse de convergence pour une capacité $C = 4 \times 10^4$. Nous pouvons observer 4 types distincts de comportement, en fonction de l'ordonnanceur (BA ou DA) et de l'égoïsme (SDM ou CDM / IDM). Nous pouvons aussi voir que l'hétérogénéité permet d'échapper à des problèmes de blocage potentiels vus pour le scénario homogène lorsqu'un ordonnanceur *BA* est considéré (Figure 4b).

Cependant, nous pouvons voir qu'un comportement égoïste rend toujours une solution sous-optimale comparé à un comportement non-entièrement égoïste pour les deux ordonnanceurs. En fait, une analyse plus approfondie des algorithmes basés sur BA montre que les algorithmes non-entièrement égoïstes peuvent plus facilement échanger des utilités de confort contre des utilités vitaux aux rounds 2 et 3. En outre, nous pouvons aussi voir que *DA* permet une amélioration de la solution plus rapide que *BA*. La Figure 9 examine les temps de résolution relatifs à *GM*. Nous pouvons avoir une analyse similaire à celle faite pour le scénario homogène. En effet, l'optimisation sur 2 maisons (*BA*) a un temps de résolution plus rapide par round comparé à une optimisation sur 3 maisons (*DA*). En outre, l'égoïsme contribue à une augmentation significative du temps nécessaire pour finir un round. Un phénomène notable dans Figure 9 est l'«explosion» de la durée de résolution relative pour des capacités élevées. Nous ne disposons pas d'explication claire pour ce comportement.



(a) Utilité vitale (b) Utilité de confort

FIGURE 8 – Utilité relative en fonction du nombre de rounds pour une capacité de 40000 Watts (cas hétérogène)

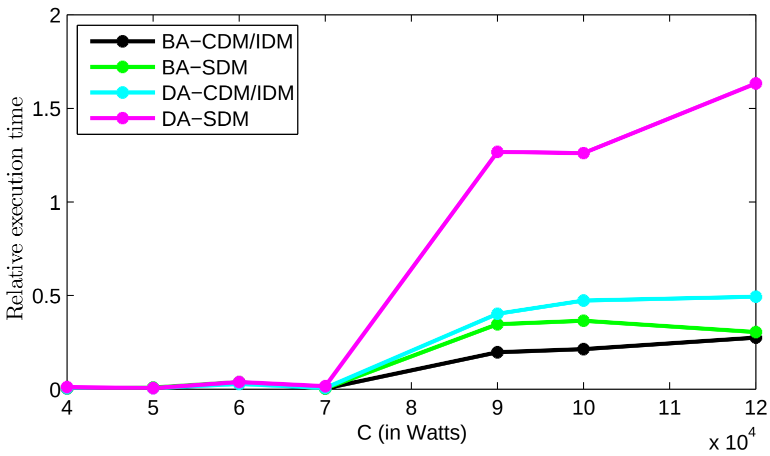


FIGURE 9 – Temps de calcul relatif (cas hétérogène)

6 Conclusion

Dans cette thèse, nous proposons et analysons des solutions de GAD avancées qui peuvent profiter de la flexibilité d'équipements hétérogènes afin d'atteindre des objectifs de forme de la courbe de charge. En effet, la flexibilité fournie par les ressources participant aux solutions de GAD est un atout majeur pour le maintien de l'équilibre entre la production et la consommation particulièrement pendant les périodes de manque de puissance disponible. Dans la Section 1 (correspondant au Chapitre 2 du mémoire), nous donnons un aperçu des approches qui permettent de répondre à ces objectifs et nous discutons de la nécessité d'une optimisation jointe des ressources hétérogènes.

Prendre des décisions au niveau d'un appareil générique est difficil : elle doit dépendre du profil énergétique de l'équipement et l'impact de la modification de ce profil sur la qualité d'expérience perçue par l'usager. Dans la section 1 Chapitre 3 du mémoire), nous nous basons sur les efforts de classification existants pour définir un nouveau framework générique permettant de décrire la flexibilité de tout équipement. Ce framework permet de construire des fonctions d'utilité pour les équipements contrôlés. Ces fonctions sont utilisées pour définir les systèmes de contrôle proposées dans cette thèse. Nous montrons que le modèle proposé pour les fonctions d'utilité est riche et assez précis pour représenter des équipements servant des fins très différentes. Quant à l'usager, elle peut attribuer à chaque équipement dans sa maison un niveau de criticité (par exemple, vital et confort) et une priorité (pour un même criticité). Cela permet à différents usagers d'avoir des préférences différentes pour le même équipement. Nous proposons également un moyen pour traiter des usagers ayant des niveaux de services différents.

La Section 1 évalue la performance de mécanismes basés sur le modèle d'utilité proposé (Chapitres 4, 5 et 6 du mémoire). Ces mécanismes agissent à différents degrés de granularité des données et de connaissance des contraintes du réseau. Ils sont de trois familles : mécanismes centralisés, partiellement distribués et distribués, agissant à différentes échelles de temps (à savoir, en avance ou en temps réel).

Nos résultats exposés dans la Section 1 peuvent être vus comme suit : les systèmes centralisés sont les meilleurs pour la performance mais sont complexes. Leur complexité peut être atténue en utilisant des heuristiques gloutonnes soigneusement conçues. Cependant, les préoccupations soulevées liées à la divulgation des informations privées des usagers peuvent empêcher leur adoption. La confidentialité peut être entièrement ou partiellement abordée en envisageant des systèmes de contrôle hiérarchique à communication unidirectionnel ou bidirectionnel au prix d'une performance plus faible. Enfin, pour nos cas d'usage considérés, il semble que les systèmes distribués peuvent fournir des performances quasi-optimales tout en préservant la vie privée et en éliminant la nécessité d'une entité de contrôle centrale.

Les contributions de cette thèse sont une première étape pour la conception de solutions de GAD efficaces capables d'intégrer des ressources hétérogènes. Nous décrivons brièvement des perspectives potentielles :

Contraintes physiques locales

Dans cette thèse, nous nous concentrons sur l'évaluation de l'impact sur la performance de plusieurs types de systèmes de contrôle en supposant des contraintes simples imposées par le réseau électrique. Ces contraintes sont formalisées en termes de la forme de la courbe de consommation globale désirée. Toutefois, il pourrait être souhaitable de profiter de l'emplacement des équipements afin de fournir une réponse plus efficace en fonction des contraintes locales du réseau physique.

La valeur de la GAD et les incitations nécessaires

Nous avons proposé des solutions de GAD qui produisent des décisions en prenant en compte la qualité de l'expérience des usagers. La haute performance du contrôle augmentera son attractivité. Toutefois, des incitations supplémentaires doivent être fournies aux usagers pour les inciter à participer à la GAD. Cela nécessite l'élaboration d'indicateurs clés de performance qui permettent de valoriser le service effectivement fourni par les usagers.

La GAD et l'incertitude

Une approche déterministe est la première étape vers la conception de systèmes qui peuvent prendre en compte l'incertitude des paramètres du problème. En effet, les algorithmes déterministes peuvent être utilisés pour résoudre des problèmes d'optimisation stochastique. Par exemple, la technique de décomposition en scénarios permet de dériver des problèmes déterministes dont les paramètres sont de réalisations possibles des paramètres incertains. La gestion de l'incertitude peut également être faite par un recalcul des variables de décisions lorsque les paramètres dévient significativement de leur valeur attendue.

Manipulation et sincérité

Dans cette thèse, nous supposons des usagers sincères et une acquisition de données correcte. Des moyens de détection de comportement malfaisant sont nécessaires. Dans notre modèle proposé, les valeurs de plusieurs paramètres sont nécessaires pour produire les décisions de contrôle adéquates. Ainsi, la conception d'algorithmes de contrôle doit permettre de vérifier la sincérité des usagers sans induire des problèmes de confidentialité.

Contents

List of Figures

Chapter 1

Introduction

The power grid is a complex system in charge of linking generation sources and energy consumers. Its key challenge is to continuously maintain a balance between generated power and demand.

Traditionally, balancing is performed on the side of production: power generation follows the load curve. This implies strong constraints on the power generation mix, like a peak-size dimensioning and the capability to match fast demand variations. Today, these constraints are met by mixing adapted types of generation sources: high-power, slow-response (e.g. nuclear plants); low-power, high-response (hydroelectricity); intermediate (coal power station). Some of these generators are very costly in terms of maintenance and/or environmental impact.

Present trends suggest that this precarious balance will become harder to maintain: on the one hand, user consumption has no reason to become less "peaky"; on the other hand, additional environmental targets are coming into picture, leading to the integration of more and more renewable energy (e.g. solar, wind) to the mix[80]. In Europe, the "2020 climate & energy package"[1] was issued to set the so-called "20-20-20" energy targets for 2020: 20% energy consumption from renewable sources; 20% reduction in primary energy use and 20% reduction of greenhouse gases below 1990 levels. Similarly, in the United States of America (USA), the "25 x '25" Renewable Energy Initiative targets getting 25% of energy needs from renewable energy by 2025[2].

In this new challenging ecosystem, flexibility and fine-grained monitoring are key needs to lower system costs and optimize its operation. As a response to these needs and to ensure grid stability, the Smart Grid concept emerged.

Smart Grid is a vision in which information, communication and control meet to attain a sustainable, cost effective and secure power system[91]. As a matter of fact, a Smart(er) Grid is capable of taking advantage of advances in Information and Communication Technologies (ICT) and the deployment of Internet of Things (IoT) enabled objects like smart appliances and sensors, to better manage and dynamically optimize its operations.

As generation is becoming less controllable (wind and sun cannot be activated on demand), a promising way for Smart Grids to secure grid operation and prevent blackouts caused by imbalances is to gain some control over user consumption. In fact, by using smart metering and the deployment of an Advanced Metering Infrastructure (AMI) allowing two-way communication between utilities and end users, it becomes possible to have the fine-grained information required to produce high quality control of power flows.

[1] http://ec.europa.eu/clima/policies/strategies/2020/
[2] http://www.25x25.org/

1 Motivation

To maintain the generation-consumption balance in a Smart Grid environment, two main solutions can be envisioned; energy storage or demand shaping. These solutions can help to cope with peaks without requiring to increase generation capacity and hence system costs.

Energy storage technologies can provide the best solution to compensate imbalances by supplementing production during high consumption periods and storing excess in production for later use. However, apart from pumped-storage hydroelectricity, they are still expensive. In addition, their return on investment of batteries highly depends on the savings made by shaping demand through charging and discharging them. Indeed, conclusions on their profitability highly depends on energy pricing schemes supposed (see [66] and [19]).

An alternative solution would be in shaping demand. Indeed, balance can be maintained by acting on demand shape to reach generation capacity goals related to actual physical limitations or cost reduction. In this context, acting on demand is an important asset that can help the grid thanks to:

- its high responsiveness compared to generators that require startup delays. For example, it is possible to reduce or increase demand instantaneously by exploiting the inertia of some loads (e.g., heating can be used as a virtual battery).
- its capacity to solve local spatial grid contingencies because of the distributed flexibility on the network (compared to traditional bulk generators connected to the transmission network of the grid).

Services that are designed to change the shape of demand curve in response to grid requirements can be categorized into Demand Response (DR) or Demand Side Management (DSM) services. While DSM refer to services targeting energy efficiency and overall load reduction (see [10]), DR relies on demand flexibility and the change in energy use at specific times. Two types of DR solutions can be envisioned to reach demand shape goals: inducing desired change or directly controlling demand. For the first type of DR, users consumption behavior can be influenced through dynamic changes in electricity price over time. For the second one, users consumption is directly controlled while providing them with incentive payments to accept such an intrusive control. The main advantage of the second type compared to the first one is the strict and assured guaranties that it can provide on the control outcome.

Nowadays, advanced DR solutions are conceivable thanks to technological advances in complementary fields like communications, control, mobile networks, cloud computing and Big Data analytics. Indeed, the wide deployment of sensing and control components is motivated by falling technology costs. Thus, home automation technologies are becoming more adopted as key enablers for a better management of energy use and efficiency. Today, smart plugs and smart appliances (e.g., smart fridges, smart thermostats) are being widely commercialized. Moreover, standards and models are being developed for smart grid architecture and IoT for the energy sector (e.g., efforts by AllSeen Alliance[3] and Home Plug[4]). An example of such standards is the Smart Energy Profile (SEP) Application Protocol Standard[5]. These efforts' goal is to ease automation of processes and allow interoperability between physical components of the sensing, communication and control infrastructure and between the different functional layers (called interoperability layers in the Smart Grid Architecture Model (SGAM) Framework[6]).

Although use cases for DR are becoming more elaborate and smart components' technology is becoming mature[46], efficient technical DR solutions are still needed. This requires the definition

[3] https://allseenalliance.org/
[4] http://www.homeplug.org/
[5] https://standards.ieee.org/develop/wg/SEP2.html
[6] ftp://ftp.cencenelec.eu/EN/EuropeanStandardization/HotTopics/SmartGrids/SGCG_Methodology_SGAMUserManual.pdf

of a control architecture. Under this architecture, interaction between entities that take part in the solution along with the distribution of intelligence among these entities need to be specified. Then, based on the specific architecture, algorithms need to be designed. In addition, DR solutions need to be able to take advantage of heterogeneity of participating resources that are seen as an asset. Indeed, heterogeneity can provide different degrees of freedom that the control can exploit. Most importantly, solutions need to be developed in view of user acceptance and adoption. Addressing these needs is the main focus of this thesis.

2 Thesis contributions and outline

As discussed earlier, two main strategies can overcome the limitations of traditional power grid, namely storing energy at various points of the network, or shaping demand through the deployment of DR solutions. However, technologies are still under development to make batteries profitable. This is why in this thesis, we rather focus on DR solutions targeting appliances flexibility. More precisely, we consider mechanisms based on direct control of users loads. This choice is motivated by the high value of strict control under user behavior uncertainty.

To design control mechanisms targeting energy efficiency and users' satisfaction, we define a generic approach for defining load's flexibility and the impact of control decisions on users. The proposed framework does not provide any restrictions on the type of controlled appliances nor on the granularity of control decisions. This enable heterogeneous loads to be entitled to joint control.

Based on the framework, we instantiate research operation problems. These problems represent mechanisms that differ in the data that is made available for control purposes and the distribution of decision making among entities involved in the control. Indeed, we examine three types of control architectures, namely centralized, partially distributed and fully distributed solutions. We evaluate the limitations of each proposal and we discuss their efficiency from users point of view by conducting numerical analysis on simple but representative use cases. We also analyze scalability, complexity and security (i.e., privacy and anonymity).

The contributions of this thesis are structured following Figure 1.1 that presents the outline of the present document. They are explained as follows:

- In **Chapter 2**, we give an overview of existing DR solutions. First, we situate DR in the existing power grid ecosystem to define their main value along with their control goals. We present the different types of DR that allow us to reach these goals and motivate our direct load control approach. Finally, we survey efforts targeting the proposal of solutions that can support the control of heterogeneous loads.

- In **Chapter 3**, we introduce our system architectural framework. The system is supposed to be hierarchical to model different functional groups. We detail possible interactions between entities that make up this architecture. This will allow us to define the different control architectures of the next chapters. We present our approach to deal with the control of heterogeneous appliances based on utility functions definition. Our proposal is discussed in view of related proposed models in the literature. It also provides means to support users that have different quality of service requirements.

- In **Chapter 4**, we propose centralized solutions based on the model proposed in Chapter 3. These solutions are capable based on fine grained dynamic information to take fine grained decision on the level of appliances to reach desired goals. We design mechanisms to provide different ways of dealing with fairness and to support ahead-of-time and real-time decision times. Proposed mechanisms are compared in view of tractability and performance.

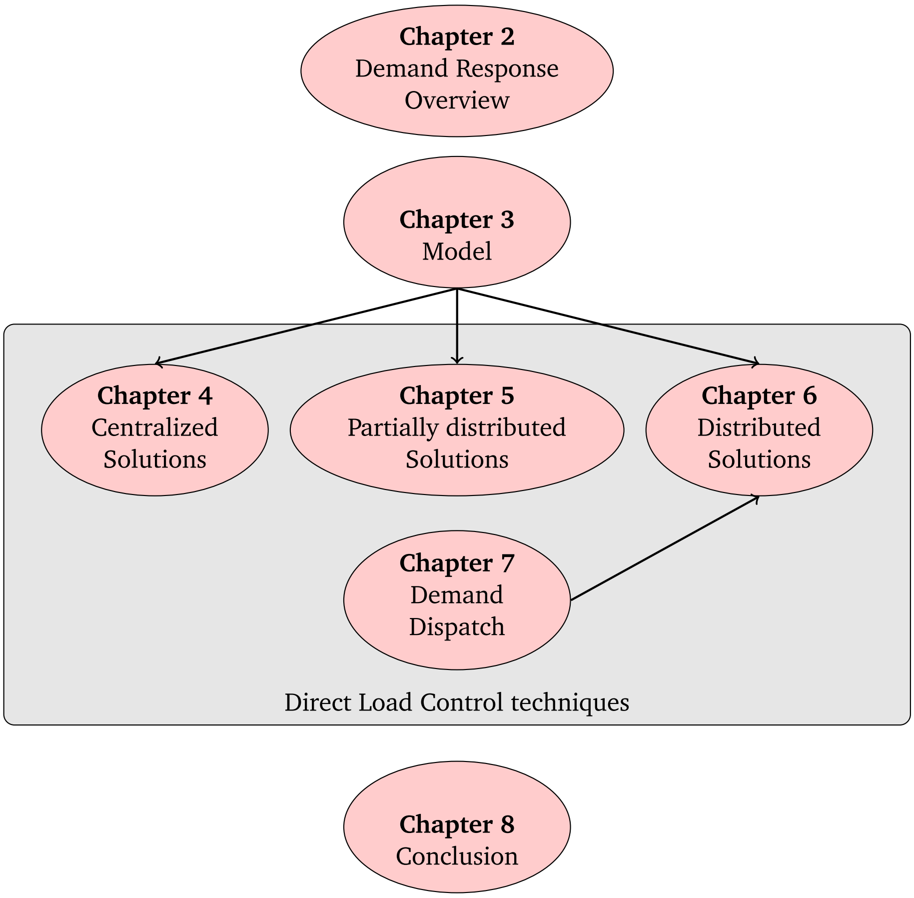


Figure 1.1 – Chapters overview

- In **Chapter 5**, we address partially distributed control schemes. These schemes allow to limit information exchange compared to the centralized approach by supposing intermediary aggregation points between appliances and the entity deciding the global control goal. We propose two types of schemes belonging to this family, namely schemes based on one-way communication or based on two-way communication. These schemes are evaluated by comparing their performance to the globally optimum decision. The proposed two-way communication scheme is also analyzed in terms of possible communication limitations.

- In **Chapter 6**, we introduce distributed control schemes. These schemes avoid the need for a coordinator introduced by a two-way communication mechanism while preserving desired privacy feature. This is done by allowing distributed optimization among entities at the same hierarchical level. Our proposals suppose different modes and levels of cooperation between entities. They are evaluated under several scenarios.

- In **Chapter 7**, we present the results of a distributed control mechanism. In this case, we suppose the control of a large population of appliances. These appliances have a local intelligence that enables them to react accordingly. We study the efficiency of proposed control schemes in view of system-wide cost.

- In **Chapter 8**, we conclude the thesis by providing a summary of results and contributions. We also provide future research perspectives.

Chapter 2

Demand Response: Overview

In this chapter, we survey efforts devoted for proposing Demand Response solutions. DR has a wide spectrum. Indeed, DR solutions can be differentiated based on their objectives and their type. The objectives of DR will define the required responsiveness of the implemented mechanism. For instance, if DR needs to provide a service (e.g., load reduction) that should be delivered in matters of minutes to the grid, the time for computing control outcome need to be adapted. Indeed, fast response require strict limitations on complexity. The type of control defines the impact on power consumers (e.g., financial, comfort related). Depending on whether the consumers are industrial, commercial or residential, expectations from DR may be different. To fix a scope for our study, in this thesis, we focalize on DR targeting residential consumers.

First, we start by presenting the history of DR development in Section 1. In Section 2, we present an overview of traditional and new trends related to power grid control. In Section 3, we introduce the actors, constraints and objectives of DR with a special focus on the time-scale of the control and the context. Then, we present the different types of DR in Section 4. In Section 5, we introduce effort dedicated for the control of heterogeneous resources. Finally, we conclude on the positioning of the contributions of this thesis in the context of Demand Response solutions.

1 History

DR and DSM became more popular with the advent of Smart Grids. However, they are not new concepts. In France, to maintaining grid stability, controlling demand started decades ago and was mainly tied to the evolution of the generation mix. A detailed history of the electricity sector of France can be found in [18]. For short, the first forms of DR appeared in the 1950s. During this period, most demand came from the industrial sector (72% in 1951). Interruption contracts were made with metalworking industries to assist the extension of the power grid and installation of hydro-power in addition to existing coal-fired power plants. These contracts with low energy price were offered to these industrial consumers in exchange for load shedding at peak demand times. Later, due to energy crisis (1970s) and the launching of “blue” electricity meters (1963), electricity consumption of the residential sector increased significantly (consumption in the 1970s was five times higher than that of 1950s). In addition, nuclear generation was introduced (1965). This generation technology has limited ramping capability: slow possible variation of power output. This made demand peaks management more important which motivated new services targeting load flattening. This lead to services offering different tariff schemes for electricity targeting residential consumers. So, a Time Of

Use (TOU) tariff called “heures pleines/heures creuses” appeared in 1965. This tariff scheme offers two prices of energy: one for low consumption hours of the day (i.e., “ heures creuses”) and a standard rate for normal hours (i.e., “ heures pleines”). It was mainly targeted to encourage the use of electric water heaters at night. With the expanded use of bi-energy (electric and fuel oil powered) space heaters, a new offer appeared that is named “Effacement Jour de Pointe” [1] (EJP) in 1982. This schemes offers a low electricity rate all over the year and very high rates for certain hours of high peak demand. Indeed, this scheme is advantageous for users that have the bi-energy heaters since they can easily change the type of energy used for heating based on electricity prices. Later, other form of DR based on interrupting loads became popular in France. For instance, distributed load shedding (“effacement diffus” in French) appeared in the years 2000 (see [36]). Its main principle consists of the possibility of shedding an aggregation of demand generated by consumers interconnected to the distribution network. Depending on the country, DR and DSM may have a different evolution timeline; a brief history of DR in the USA can be found in [17]).

Even though the concept of DR existed, it was very limited because of regulatory and technical barriers. Indeed, telemetering was limited and consumption data acquisition was based on meter reading at each physical location. In addition, markets rules did not include clauses that allow direct DR participation. With the development of Information and Communication Technologies (ICT) and the advent of smart grids, the scope of DR solutions has extended. Indeed, it became technologically possible to design an automated system capable of supporting DR services. Systems can be designed to provide, access and process relevant dynamic information and take required actions when necessary.

2 From load following to load shaping

A key rule in maintaining power system stability is to balance generation and consumption at all times. This is traditionally done by controlling generation sources that should have a power output that matches demand shape. However, with the introduction of renewable energy, generation is becoming less controllable. This gave rise to Demand Response as a way of shaping demand to overcome generation flexibility shortage. We now give an overview of traditional system control, new trends with the deployment of renewable energy sources and possible ways to deal with demand shape through DR.

2.1 Traditional generation sources control

As stated previously, in traditional power systems, generation should be controlled to always match demand. This is denoted by *load following*. In order to have enough available generation sources capable of following the load curve, system planning is done on different timescales. These timescales are related with existing markets that determine unit commitment for each production source. These markets are designed to adjust generation in order to fulfill predicted demand or correct demand prediction errors. Indeed, some demands can be predicted with high certainty a long time in advance which allows an early reservation of generators. These demands correspond to base load and usually take advantage of (somewhat cheap) slow ramping sources like nuclear.

For more uncertain demands, power generation required to fulfill them can be decided on markets of faster timescales namely wholesale spot markets and Ancillary Services (A/S) market.

For wholesale spot markets, each production source will know how much it should generate at each hour of the next day. These markets are usually designed to decide the price of energy depending

[1] https://particulier.edf.fr/fr/accueil/facture-et-contrat/contrat/consulter-les-jours-ejp-et-tempo/option-ejp/details.html

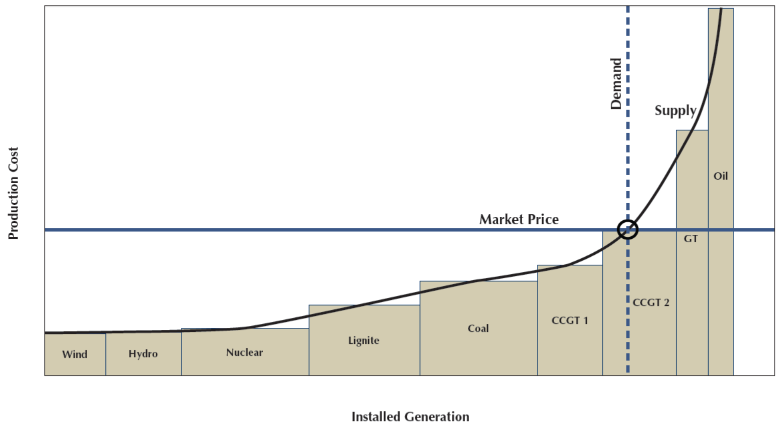


Figure 2.1 – Supply stack (Source: [96])

on the marginal cost function of generation. Let us suppose hourly markets. The cost function shape depends particularly on the generation mix available at a given hour on the grid. It usually grows exponentially with the amount of energy that need to be generated (see Figure 2.1 that represents the mix for a given hour). In addition to generation mix, this cost function will depend on production capacity and ramping capability of generation sources supposing the supply outcome at the previous hour. So for a certain hour, settlement price for energy is equal to the marginal cost at the point where demand curve and supply curve intersect. The amount of energy that should be generated is given by the abscissa of the intersection point. Most generation sources of average cost (e.g., nuclear) have slow ramping capability. So, when a peak in demand occurs costly flexible generators (e.g., gas and oil power plants) need to be used. This will yield in a higher price for energy.

Wholesale energy markets can, however, be insufficient to maintain the balance since demand prediction is never perfect and unpredicted events may occur (e.g., bad control of a generator output). This requires sources that can react on a faster time scales to imbalances (in the order of minutes to seconds). Services targeting to solve fast imbalances are referred to by A/S. There is mainly three types of A/S, namely *primary*, *secondary* and *tertiary* regulation [83]. Imbalances in generation and consumption will make the frequency of the grid deviate from its nominal value (i.e., 50 Hz or 60Hz depending on the country). This nominal value corresponds to balanced generation and consumption. When an imbalance occurs, primary regulation is used to stabilize grid frequency fluctuations. This should be done in matters of seconds (e.g., in less than 30 seconds in France). Following primary reserves, secondary and tertiary reserves are consecutively used in order to bring back frequency to its nominal value (in matters of minutes). This is done by producing more or reducing generation in a way that equates consumption and production on the grid. Indeed, secondary reserves are immediately automatically dispatched. Tertiary are called for when secondary reserves are insufficient or to preserve flexible capacity of sources participating to secondary regulation. Generation sources providing A/S need to have required capacity and enough ramping flexibility to follow adequately the load curve.

2.2 New renewable intermittent sources

Maintaining balanced generation and demand is made more complex with the introduction of renewable energy sources and the willingness of governments to go towards clean energy.

With the increasing deployment of renewable energy sources and particularly highly volatile ones like wind turbines and PhotoVoltaics (PV) cells, flexible resources that can help maintaining the bal-

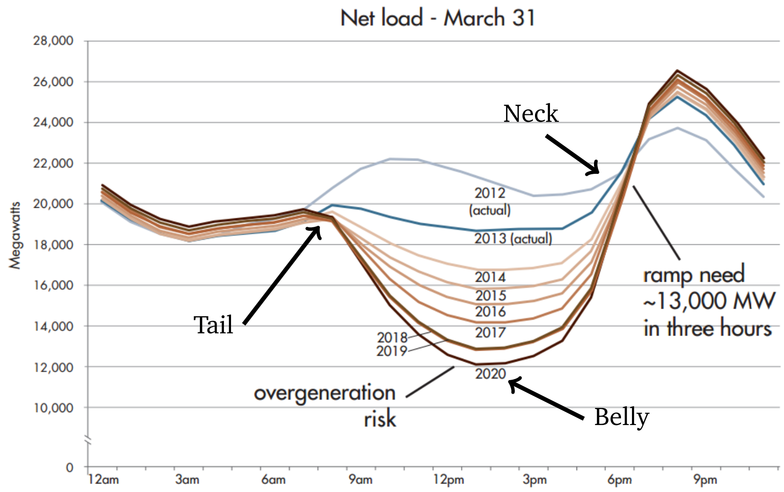


Figure 2.2 – Duck shape (Source: CAISO)

ance between generation and consumption, is becoming more needed. One example of potential effects of high penetration of renewable energy sources can be illustrated by the "duck shape" (see Figure 2.2 from California Independent System Operator (CAISO)) of the net load curve that may appear in certain times of the year[2]. Net load is the difference between forecasted load and expected generation from renewable energy sources. It corresponds to the energy that should be produced by classical generation sources in order to maintain a consumption and generation balance. The observed shape can be explained by taking the example of solar generation. Indeed, solar generation will be concentrated during some sunny hours of the day. During these hours, the net load curve will have very low values which produces the shape of a belly. The belly will be more pronounced with the increased penetration of the energy source (i.e., solar in this case) and may sometimes result in negative values of the net load (i.e., over generation). In this case, enough flexible resources should be available particularly to cope with steep ramps (e.g., on the tail and the neck of the duck).

2.3 Main objectives of Demand Response

As opposed to having generation sources that should produce a supply curve that matches demand, DR enable to invert objectives. Indeed, it allows to shape the demand in order to reach a desired generation curve. Figure 2.3 illustrates typical objectives of DR [28]. To shape load curve, six typical operations can be implemented that are:

- *Peak clipping:* reduces demand peak when consumption exceeds a desired supply limit. This limit can be related to cost reduction objectives or to physical limitations (e.g., generation shortage, infrastructure constraints).
- *Valley filling:* increase the consumption during off-peak hours. This is particularly interesting to increase the usage of slow ramping generation sources while having generations means that can cope with the peak. Indeed, these sources are usually less costly than flexible ones.
- *Load shifting:* moves part of the demand from peak hours to off-peak hours in order to flatten the demand curve. This allows to fulfill most of the demand by slow ramping generation sources.

[2] https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf

- *Strategic conservation:* lowers energy consumption during certain hours without changing ramping features of load curve. This can be attractive to lower the generation output of slow ramping resources.
- *Strategic load growth:* increases energy consumption during certain hours without changing ramping features of load curve. This can be desired when excess in generation occurs. For instance, it can help lowering the level of water in dams or to consume the excess of generation from renewable energy sources.
- *Flexible load shaping:* changes the load curve to match a desired shape. For instance, it can be desired to have a load curve that matches generation from renewable energy sources.

Next, we will detail how these desired load shape objectives can be achieved.

3 Objectives and approaches

DR development has been tied to each country's special regulation, power grid operations and markets' structure. However, regardless of the possible country specific characteristics, several common ground objectives can be identified from DR mechanisms. These objectives usually depend on the scope and the time scale of the DR control.

In this section, we will break down the main contexts and corresponding control timescales in which demand response is beneficial. We also discuss typical control objectives for actors involved in DR. This provides background material that will allow us to situate the contribution of this thesis by giving intuition into possible real-world applications of the control schemes we will propose in the next chapters.

We start by giving an overview of the main actors implicated in the DR ecosystem. Then we present the different categories of DR that will define how and by whom it can be used as a system resource. Then, we elaborate on control objectives by approaching DR benefit from a control time scale point of view. Considered time scales are defined with respect to existing markets. We next present evolution of markets to support DR. Finally, we discuss objectives and constraints related to physical power grid operations.

3.1 Main actors

In the DR ecosystem, actors can be split into customers, service providers and service consumers [90]. An entity, however, can combine several roles.

Customers refer to energy consumers that participate to the DR service and offer consumption flexibility. Flexibility is exploited by DR in order to reach certain goal discussed in the following sections (see Figure 2.3). As stated previously, we will be focusing on residential energy consumers. Their participation can be mandatory or optional depending on the DR service provided (see Section 4).

The DR provider can be a third party company or any energy market participant like utility companies, suppliers or system operators. Utility companies and suppliers are the entities responsible for buying energy required by their clients on energy markets to fulfill their consumption needs. So, they decide the pricing schemes for energy for their clients. In the rest, we will refer to the DR provider by "aggregator" since it aggregates demand flexibility of its DR customers.

DR consumers are entities that buy DR services (for their provided flexibility) to balance generation and consumption in their domains. Indeed, maintaining grid stability requires generation and demand to be balanced at all times. However, with the deregulation of energy markets, a large number of actors became involved in market transactions which makes maintaining the balance more complex. To cope with the introduced complexity, each market participant is responsible of a domain that depends on its role. So each participant should strive to balance consumption and production in his domain. A domain can be made of physical and/or economic flows. Physical flows correspond to metered power flows on the power grid whereas economic flows refer to financial transactions. Let's consider the example of a supplier. A supplier can be a DR consumer. Indeed, it should balance the consumption of its clients (i.e. a physical flow based on metered data) and production on its domain. In this case, production will depend on the contracts attesting the amount of energy he bought on energy markets (i.e., an economic flow). DR consumers are usually referred to by balancing authorities. To enforce balance, penalties are imposed if a balancing authority fails to equate input (consumption) and output (production) flows on its domain.

We will see how actors can trade DR resources in Section 3.2. The timescale at which these resources can be needed will be detailed in Section 3.3.

3.2 Implicit versus explicit services

Defining the purpose DR may serve will depend on how this resource can be utilized in the system. This may constrain the timescale of DR schemes (see Section 3.3) as well as markets and grid operation optimization to which it may participate. This also constrain the type of solutions that can be implemented (see Section 4).

Let us suppose a market that will dictate the timescale and objectives of DR solutions (e.g., day-ahead market where DR can help bring down the energy price). Depending on the DR provider and consumer, DR services can be seen as *implicit* or *explicit* with respect to the targeted market (see[32]).

It is explicit when the DR provider propose it as a product on the market. In this case, market participants can buy it similarly to a generation resource. DR explicit participation to a market usually require the existence of regulatory framework. These frameworks constrain and set the rules for DR participation. In [99], examples of DR products are given from Pennsylvania-New Jersey-Maryland Interconnection (PJM)[3] and New York Independent System Operator (NYISO)[4].

DR can implicitly participate to a market if the scheme affects the outcome of the market without being an explicit product of this market. This is the case when the supplier is the DR service provider and consumer. In this case, the supplier can change its buying strategies on the market depending on DR opportunities from his clients participating to DR. This is usually the first and simplest forms of DR that do not require the existence of regulations.

3.3 Control timescale

Requirements for DR control schemes are two-fold: those imposed by DR customers (i.e., end users) and those imposed by DR consumers (e.g., utility company, grid operators).

For end users, acceptance of DR programs depend on the perceived quality of experience, financial incentives (reduce their bills) and privacy considerations. Some users may also have environmental objectives (help reduce usage of polluting generation sources and promote renewable energy sources).

While DR customers objectives are usually similar, the purpose of DR services and their value depend on the time scale of the control. In this context, control time scale expresses the duration between the time at which the control is decided (load commitment) and the time at which DR-related consumption change should become effective (the service is delivered).

[3] http://www.pjm.com/
[4] http://www.nyiso.com/

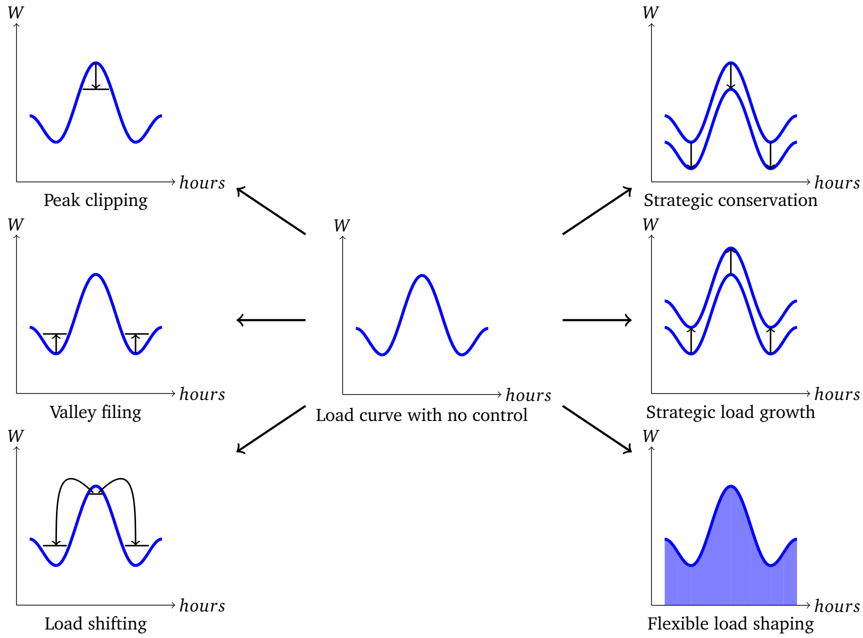


Figure 2.3 – Demand Response typical objectives

Based on existing markets, different categories of DR can be identified. Theses categories can be roughly grouped in "slow" DR, "real-time" DR and "fast" DR mechanisms ([77]), depending on the time scale.

In this thesis, we will propose mechanisms that can be implemented by a DR provider. Proposed mechanisms' timescales correspond to the so-called "slow" DR and "real-time" DR. So, our proposals will target answering requirements relevant to these timescales.

Slow DR

Slow DR mechanisms are usually designed in order to optimize costs on day-ahead wholesale markets. On these markets, price for energy settlement for a certain hour of the next day usually depends on the marginal cost function of generation (Refer to discussion in Section 2.1). To lower energy price, demand curve need to be shifted by lowering consumption. This is mostly desired when a peak in demand curve occurs that produces a high price. Hence, for most proposed "slow" DR mechanisms, utilities or aggregators' goal is to shave the peak. This can be done by minimizing peak-to-average ratio of the demand curve which allows to have a lower settlement price for energy (e..g, [60],[69]). To smoothen and flatten the load curve, typical DR objectives consist in techniques like peak clipping (i.e., reduce energy consumption), valley filling and shifting demand from peak hours to off-peak hours (see the left hand side graphs in Figure 2.3).

These goals, however, need to be evolved when introducing renewable intermittent energy sources in such markets. Actually, since these sources usually have a zero marginal cost, the supply curve may be shifted to the right depending on the amount of generated power by these sources at a given hour (see Figure 2.1). So, in this case, it is more adequate to minimize peak-to-average ratio of the residual from demand curve after subtracting renewable energy generation need to be minimized as a way of minimizing price settlement. In addition, several studies like [78] show the need for an evolution in the market design from marginal costs in order to support the growth of renewable energy sources on the grid.

We note that slow DR can also refer to intra-day hourly wholesale markets. In this case, suppliers and balancing authorities resort to exchanging bilateral (e.g., Block Exchange Notification service in France[5]) in order to correct imbalances to reduce penalties (see Section 3.1). Implicit "slow" DR can in this case be another possible choice for suppliers (see Section 3.2).

Real-time DR

Real-time DR are usually designed in view of real time intra-day markets. These markets usually set prices an hour-ahead or 15-minutes ahead. In some countries, real-time markets can have faster time scales up to 5 minutes (e.g. in PJM[6] in the USA and Independent Electricity System Operator (IESO)[7]) in Canada.

Based on the category of DR (see Section 3.2), two cases can be identified:

- For implicit DR, proposed schemes are related to supplier balancing needs. In this case, suppliers may implement real-time DR in order to correct their demand prediction errors.

- For explicit DR, two scenarios can be considered depending on the market and its participants. First, if the market is related to providing A/S to the grid, the DR consumer is the system operator (e.g., Transmission System Operator (TSO), Independent System Operator (ISO) or Regional System Operator (RSO)) or a balancing authority of a physical area. The goal of A/S is to correct prediction errors and restore grid frequency to its nominal value (corresponding to production being equal to consumption). This correspond to the so called tertiary and secondary reserves ([83]). In this case, participating generators and DR resources have to follow a signal sent by the system operator. The timescale of the signal is usually in the order of minutes. The control signal is computed from the Area Control Error (ACE)[8] that is a combination of frequency deviation from the nominal value and prediction errors. Secondly, if the market is a real-time energy market that follows similar rules of energy markets discussed for slow DR, objectives and the purpose from DR will be similar for both markets. However, in this case, DR control schemes need to have a faster reactivity.

In addition to market cost optimization, the accuracy of control for this time scale becomes a more important issue. Indeed, if control is inaccurate and acts against balancing needs, the whole energy system may be jeopardized as it will require additional resources to correct generated imbalances at a faster time scale. In the USA, the North American Electric Reliability Corporation (NERC) imposes bounds on the error signal (i.e., ACE) values in a given area defined by Control Performance Standard 1 and 2 (CPS1 and CPS2). So, balancing authorities and system operators have reliability criteria, in addition to economic ones, to choose among available generation and DR resources. Hence, Proposed demand response schemes need to know the flexibility capacity of resources being controlled to accurately meet its commitments and goals.

Fast DR

For Fast DR, the timescale is in the order of seconds. On this time scale, the control goal is related to stabilizing the frequency of the power grid. Stabilized frequency is usually different than the desired nominal frequency value (that will be recovered by secondary and tertiary reserves). This is called primary frequency control. In this case, communication delays becomes a main concern. For

[5] http://clients.rte-france.com/lang/an/clients_producteurs/services_clients/service_neb.jsp
[6] http://www.pjm.com/markets-and-operations/energy.aspx
[7] http://www.ieso.ca/Pages/Participate/Markets-and-Programs/Real-time-Energy-Market.aspx
[8] https://www.naesb.org/pdf2/weq_bklet_011505_ace_mc.pdf

generation sources, Turbine governors are the main source used in primary frequency response. For these sources, decisions are usually taken in a fully decentralized fashion based on local frequency metering information and some predefined control parameters.

To be capable of providing primary control, DR services need to be highly responsive. In addition, DR need to provide a reliable services while controlling a potentially a large number of resources. Indeed, it was shown in the master thesis [14] that synchronization problem may arise when a deterministic method similar to the one used for controlling generators is implemented for DR resources (e.g., [88]).

3.4 Evolution of energy markets

In order to promote DR mechanisms and flexible generation to support the increasing deployment of renewable intermittent energy sources, new markets have emerged and deployed in some countries. We now elaborate on some examples of new markets that we consider as the main context in which control schemes we propose can be applied.

Among these markets, we can cite capacity markets. The main goal from this market is to secure supply. Securing supply means that enough resources are available in order to ensure a balance between generation and consumption at all times. To do so, each supplier should hold enough capacity certificates depending on the power consumed by his clients. These certificates are bought from generation sources and DR aggregators and attest that the selling resource is capable of a certain maximum power production or reduction. Penalties apply when consumption is higher than acquired capacity certificates. By buying the certificates from aggregators and generators, financial incentives are provided to maintain these resources on the grid. DR participation in this case can be implicit or explicit or both (see Section 3.2) depending on the specific deployment of the capacity market. For this application, only slow and real-time DR can be provided. Indeed, fast DR do not support economic dispatch because the main goal is related to system reliability. Other markets are also proposed to specifically support micro-grids and maintain local stability (e.g., [29]).

To promote efficiency of control proposed for real-time system services, new signals and market products are proposed. As an example of real-time system services evolution, PJM proposes a decomposition of the ACE signal into two types of signals named RegA and RegD that are sent to participating resources[9]. The former is destined for resources with slow ramping capabilities whereas the latter is suitable for highly flexible sources. To enforce tracking efficiency of a signal, a pay for performance strategy is adopted. Indeed, each resource participating to regulation will be paid depending on the type of signal it is tracking (i.e., resources following the highly dynamic signal RegD will usually be paid more than the ones following the slow signal RegA) and the performance of the tracking. Tracking performance is defined by $A\ Score_A + B\ Score_D + C\ Score_P$ where A,B,C are positive constants with $A + B + C = 1$ and where:

- $Score_A$ (Accuracy): measures the correlation between the control signal and the regulation unit's response. This score is high when accuracy is high.
- $Score_D$ (Delay): measures delay defined by the time between the control signal and point of highest correlation (from Accuracy analysis). This score is high when delay is low.
- $Score_P$ (Precision): measures the difference between the areas under the signal and the response curves. This score is high for high precision.

Resources are admitted to provide system services if they pass three tests with an accuracy higher than 75%. They get disqualified if, on a 100 hour rolling average, average performance score is below 40%.

[9] http://www.pjm.com/markets-and-operations/ancillary-services.aspx

To allow and promote DR resources participation to regulation especially for tracking RegD signal, the regulation dispatch signal sent to these resources is made energy-neutral. Energy-neutrality means that, over a period of time, the dispatch signal will have an approximately equal amount of energy above and below the midpoint of the DR resource. This is important since the capacity of DR resources can be quickly depleted if the signal is constantly requiring a specific type of flexibility (i.e., increase or decrease in consumption). Indeed unlike generating sources, reducing consumption for a long period of time will significantly affect end users and increasing consumption has its limits for maintaining user comfort. So, DR resources act like a virtual battery that has a capacity limit and can be "charged" (i.e., increase the consumption) and "discharged" i.e., decrease the consumption) depending on the state of controlled loads. Similar approach is proposed by authors in [8] for tracking a signal by a collection of DR resources. Indeed, resources with a certain operation cycle duration will be able to accurately track a signal that has an adapted ramping speed. Based on this observation, authors propose to decompose the signal to track into multiple signal depending on the capacity and the tracking capability of the DR resource. We propose a scheme that builds on this work in Chapter 7.

3.5 Power grid operation constraints

Power grid operation cannot always and exclusively be managed by markets. Indeed, additional requirements related to the physical grid infrastructure may be raised. This is because markets may sometimes be insufficient to enforce the balance between generation and demand. In which case, system operator has to take required measures to minimize the impact on the power system. In addition, markets cannot capture physical constraints imposed by the power grid. Indeed, grid infrastructure (made of equipments like transformers and lines) can support limited load capacity for a secure operation.

In theses aforementioned cases when markets cannot resolve congestions and imbalances, system operator need to take required actions in order to mitigate the emergency situation and minimize the risk of accidental blackouts and cascading blackouts.

In a traditional grid operations without DR, actions taken depend on the nature of the emergency situation. Two types of situations can be differentiated:

- When crisis situation is predictable, a typical action is to schedule planned blackouts on parts of the network. Such situation may happen for instance during hot summers and cold winters.
- In case of unpredicted crisis situation, grid protection relays are triggered. This may happen when an unpredicted failure occurs.

Since the distribution network is radial, cutting parts of the network involves cutting certain branches of the distribution tree. In addition, depending on the amount of load that need to be reduced, certain consumer' classes will impacted. Indeed, when clients are interconnected to the power grid, they will be connected to the branch of their customer class. This allows to have important public services like hospitals on certain dedicated branches of high powering priority. Restoring power after interruptions is usually difficult and costly since once customers are reconnected demand will generate a peak from interconnected appliances and synchronized demand generated by certain types of appliances (like heating) that show a pay back effect dependent on the interruption duration.

In a smart grid context, DR can help mitigate such scenarios thanks to their scattered distribution on the grid which allows local control. In [37], a typology of demand response and demand side management schemes is presented along with some success stories of demand response mechanisms that prevented planned blackouts by controlling air conditioners and refrigerators. In [50], the authors introduce a family of new DR service models for the provision of electricity, in particular to face such periods of time during which it is not possible to cope with all the demand. This can be due to physical grid capacity limitations. The proposed service models allow to significantly limit the impact on quality of experience during such time periods by avoiding coarse-grained control mechanisms (like rolling blackouts). Based on the general architecture proposed in the document, available power can be distributed in a smart way, at a fine granularity, taking into consideration the relevance of each appliance for the end-users. In particular, the proposed architecture leverages the capabilities of the Internet of Things paradigm. The mechanisms we propose in the present thesis can be embedded in this general framework as an implementation of one of its key building blocks, but their scope of application is much broader (see Chapters 4, 5 and 6).

Another physical constraint of the power grid is the need for a continuous upgrade of the Transmission and Distribution (T&D) equipments in particular to cope with the evolution of the needs. With respect to that perspective, DR can alleviate the burden. Indeed, As it may lead to a more efficient use of the infrastructure capacity and optimal loading, DR can allow to differ the need for upgrades.

4 First classification

In this section, we aim at providing a state-of-the-art overview of proposed technical solutions to the objectives and requirements identified previously. We will do that based on a classification of proposed mechanisms. Indeed, DR mechanisms can be classified into two main types, namely, pricing-based DR and Direct Load Control (DLC).

Then, for each type of DR mechanism, we will discuss proposals in view of their timescale to elaborate on how they answer the requirements identified in Section 3.3. We also overview proposals targeted for dealing with privacy and fairness considerations.

Figure 2.4 summarizes the main applications of DR based on its type and control timescale. It shows load commitment timescales at which flexibility can be reserved. Indeed, depending on system optimization objectives, flexibility can be reserved years in advance with respect to the delivery time. This is mainly for generation capacity. In this case, the main objective is energy efficiency through system planning and dimensioning. DR can usually be planned on such long timescales however it will mainly consist of static strategies. We now discuss the two types of DR showing in this figure.

4.1 Pricing-based DR

Pricing-based DR aims at modifying users consumption behavior by changing the price of energy during the day. Since energy prices are usually set by the supplier or utility company, this type of services can be seen as an implicit form of DR from a market perspective (see 3.2). Another form of pricing is to provide rewards for users that change their consumption behavior. They are very similar to directly changing electricity prices but do not expose users to extreme price changes. An example of such scheme is Peak-Time Rebates (PTR) where users get payments for the “actual” load reduced determined by comparing their consumption curve with a baseline. Some demonstrator programs (e.g., [44]) showed that this type of pricing usually favors consumers that have high consumption (i.e., high energy consumption baseline) over users that control their energy use (e.g., low income homes). This is due to the fact that the baseline of a user is usually defined based on historical consumption data. This makes the outcome of this scheme vulnerable to gaming (e.g., users may consume more to increase their baseline) and results in frustration for consumers that are energy efficient.

The impact on users subscribing to pricing-based DR services will be financial. So, it’s up to the consumer to decide to which level he is willing to sacrifice comfort in order to reduce her energy bill.

We now discuss proposed pricing-based DR mechanisms and we focus on the analysis of three dimensions, namely the timescale of the control, privacy considerations and fairness.

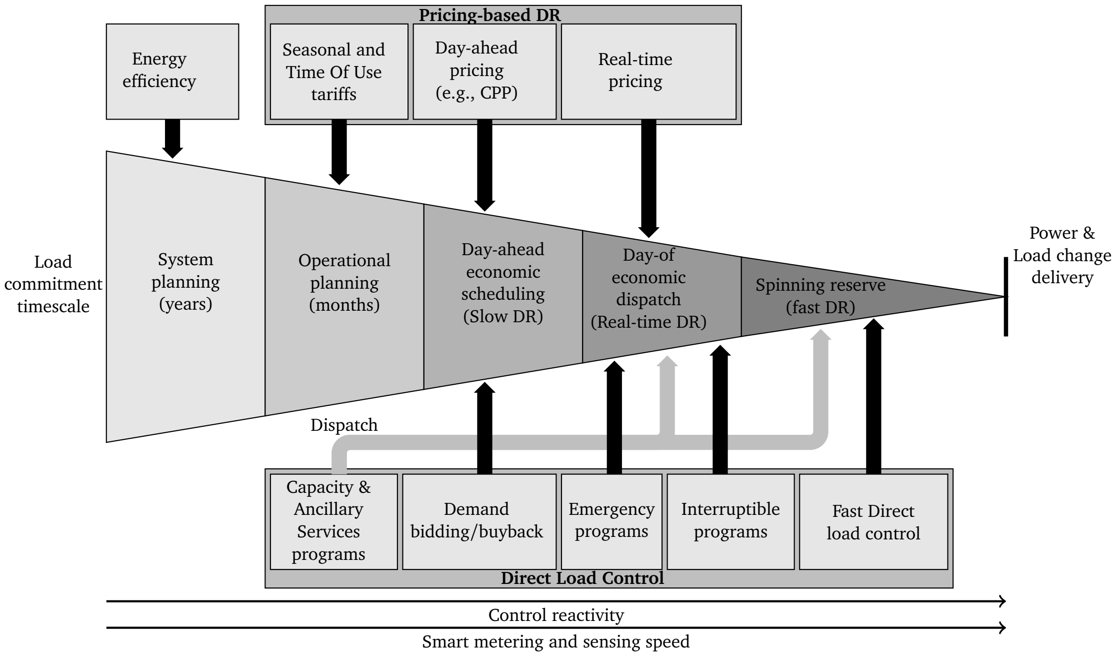


Figure 2.4 – Demand Response (based on [90],[73])

4.1.1 Timescale

Proposed schemes can either be slow or real-time. Indeed, pricing programs are not capable of providing reliable response required for fast time scales (e.g., A/S).

Several studies have been carried out to compare pricing schemes based on their time scale and dynamicity [42]. Most work capitalizes on the effectiveness of real-time approaches compared to less dynamic ones. For instance, the author in [11] shows the benefit of Real-Time Pricing (RTP) on the long-run considering hourly changing prices. In [84], authors emphasize on the required trade-off between volatility and system robustness, and efficiency. Indeed, while real-time prices are efficient from a control perspective for responsive users, it introduces system volatility due to the uncertainty in demand and generation. In addition, as suggested in [42], users tend to favor slow pricing based DR over real-time ones. Indeed, slower time-scales give a wider time visibility that allows users to better plan their electricity use. For real-time dynamic prices, users should react with the limited information they have which may result in extreme reactions and disengagement (e.g., [7]). Furthermore, real-time pricing requires the existence of an automatic intelligent system that can control home energy usage and schedule appliances to reach its full potential.

Slow DR

We distinguish between Time Of Use (TOU) electricity prices and dynamic prices. TOU electricity rates are usually set months in advance to encourage consumers to shift their demand to periods of low consumption. It is a basic evolution from providing flat prices of energy (that remain fixed on a larger time scale). Implementing this pricing scheme proved its benefit in reaching desired goal (e.g.,[20]). For DR schemes based on dynamic pricing, utility companies set hourly prices for customers to minimize its costs on the wholesale energy market.

Some proposed schemes are specifically targeting peak reduction by supposing Critical Peak Pricing (CPP) strategies. An example of this type of pricing is given by [67] in which "Inclining Block Rate (IBR)" scheme is supposed, which consists of different prices for energy for different ranges of aggregate consumption of users. We note that this proposal (and other slow DR proposals) is a Real-Time Pricing (RTP) scheme. However, in this case, real-time prices refer to time-varying dynamic prices that are fixed ahead of time [73]. The use of this naming is due to the lack of agreement on vocabulary in the energy domain.

Real-time DR

For real-time DR schemes based on pricing, energy rates are sent in real-time (i.e., an hour to 5 minutes ahead) to users.

As an example of such a scheme with an emphasize on appliance level modeling and users' satisfaction, we can cite [55] that proposes a real-time pricing scheme that takes into account the utility function of the appliances being served and show that a single price per time period can be computed for users to reach utility objective while satisfying users' objectives. In the present dissertation, we consider a similar approach with regard to the definition of utility functions. We will get back to this in Chapter 3.

4.1.2 Privacy

Pricing-based DR solutions may raise privacy concerns for residential consumers when feedback is sent to the supplier. This is the case when prices are dynamically computed based on consumption information.

One possible solution to mitigate privacy related issues is to limit the feedback to suppliers by implementing cooperative schemes between users. Indeed, if anonymity is guarantied between users, privacy can be preserved. Some work proposed cooperative schemes among users to improve their energy prices in a way that suits all (e.g., [69],[34],[68]). In [69], authors introduce a user specific day-ahead pricing scheme that charges users proportionally to their participation to the aggregate consumption for each hour of the day. In this case, users should cooperate to minimize their bills by changing the schedules of shiftable loads which will change their aggregate consumption for each hour. In [34], appliances are classified into classes that express the flexibility that they may offer (e.g., flexible level of power consumption, shifting capability). Appliances scheduling will be a result of cooperation between users to stabilize the price of electricity for each hour in a way that minimizes their respective bills. [68] proposes a pricing scheme for which users have incentives to collaborate to reduce peak-to-average ratio of their aggregate consumption which will reduce generation costs and hence their bills and enhance system performance. To do so, homes announce each their aggregate consumption in an asynchronous manner and update it depending on others announced consumption until reaching a Nash equilibrium. One drawback of the proposed scheme is the amount of information that needs to be exchanged among homes. In addition, appliances are characterized by power consumption limits during specific time periods (home management system should decide in advance in which time slots power can be positive and in which appliance is turned off) which adds constraints on shifting capabilities.

4.1.3 Fairness

Defining fairness for pricing based DR depends on the pricing scheme considered. Indeed, one can argue that providing the same price for all users given the same DR service as fair. However, user dependent prices can sometimes better capture user's effort with respect to working towards meeting power grid constraints. An example of such cases is the CPP scheme proposed in [47]. In this proposal, day ahead time-dependent pricing scheme that minimize costs for utility companies while giving incentives rewards for users to shift their loads to different time periods of the day based on their specific flexibility.

We note that, in addition to fairness, considering homogeneous prices for energy implies an implicit assumption about users react differently to prices. Actually in case of peak reduction, it is shown that homogeneous prices per time slot for all users may create artificial peaks during time periods of normally low consumption [19]. Indeed, if all users react the same way to prices, the peak can be shifted to another period of the day. User specific prices can come as a solution for such a risk where each user will receive rates based on his specific consumption habits. Provided with the prices, users can schedule their appliances in a way that minimizes their bills while maintaining comfort.

4.2 Direct Load Control

Direct Load Control (DLC) allows to have a tighter guaranties on the outcome of the proposed control compared to a pricing based approach. It is sometimes referred to by incentive based DR. Indeed, participating users will usually receive incentive payments for allowing their appliances to be controlled. In this thesis, we will be focusing on this type of DR as it offers guarantied demand flexibility.

For all control timescales, two types of control strategies can be supposed namely on a fixed horizon (i.e., fixed optimization period) or an infinite horizon (i.e., the control service runs all the time and can be called for at any time). We differentiate two types of direct control namely, compulsory and optional. For the compulsory schemes, once users decide to participate they cannot opt-out. This is usually the case when DR scheme is proposed as an explicit market product (see Section 3.2). For the case the DR service is deployed and can be called at any time, users can have the option of choosing to derogate. This type of service is called demand dispatch [15].

Some work focused on modeling the effect of DR while taking into account the characteristics of the physical network and the locations of loads (e.g., [39],[56]). While its an important step towards studying the effect of DR, a better understanding of users expectations and the effect of decisions on users is need. So, in this thesis, we do not model power flows constraints and physical infrastructure and we focus on analyzing performance in terms of the impact of decisions on users' quality of experience.

Like for pricing, we now focus on the timescale of the control, privacy and fairness.

4.2.1 Timescale

When proposing direct load control schemes, control decisions are activated instantaneously but computing schedules and control decisions can be either decided ahead of time (i.e., days-ahead and hours-ahead), in real-time (minutes-ahead) or instantaneously (in order of seconds) when a signal is sensed.

Slow DR

Slow demand response usually targets to control demands on a fixed known control period. Such schemes allows to foresee control outcome which might be attractive for users to adapt accordingly and be prepared to potential consequences (related to comfort). This type of flexibility can be used on energy markets through demand bidding and buyback. Authors in [70] propose a three steps technique to manage generation of a virtual power plant made of local generation from homes (i.e., from Micro Combined Heat and Power (microCHP) devices) considering a hierarchy of controllers. Step one consists of predicting demand for the next day locally by homes. Step two consists of a global planning to reach a desired production plan for the next day. This is done though control "price" signals (this does not mean that the approach is price-based, see below) that are sent to specific homes to change their production plan (constrained by step one) until the matching error between desired plan and aggregate production of homes is minimized. Finally step three consists of scheduling locally in real-time generation to match the one that was fixed the day before having real demand of residents (since the plan is based on prediction). Authors show the high impact of prediction errors on the performance of the proposed scheme. In [6], authors propose a tool that uses sensing information and analytics to infer user preferences and allow automatic control of shifting flexibility without requiring user's physically do the task. The goal of the scheme is to lower user's bill in case of day-ahead set prices. It is considered as a load control mechanism because the main objective is not pricing but the proposition of a framework to control load locally by an autonomic central controller. The goal of the controller is to minimize the bill and the cost of shifting demands which is supposed to be proportional to the distance between the primary schedule of appliances (modeled as chunks of energy blocks) if no control is applied and the schedule computed after applying control. Authors in [31] propose a scheme for appliances day-ahead commitment to being controlled that takes into account users' comfort requirements when a real-time pricing is supposed. Since commitment is based on a prediction of the real-time prices, slow DR scheme is followed by a real-time mechanism to adjust the scheduling. In this thesis, we propose and study DLC control schemes falling in this category.

Real-time DR

Real-time direct load control usually targets to offer ancillary system services. However, they can also participate to emergency programs to help the grid during a contingency. Interruptible programs are usually targeted for large consumers (i.e., industrial and commercial). In addition, most work on real-time DR based on residential direct control of loads at a fine appliance-level granularity has focused mainly on flexible loads like Electric Vehicles (EVs) and Thermostatically Controlled Loads (TCLs) (e.g., air conditioners, heat pumps, water heaters, and refrigerator).

Authors in [58] propose a scheduling technique that supposes real-time critical peak pricing with fixed structure. Prices are used for producing control decisions for appliances that can be shifted, depending on their allowed delay while taking into account users discomfort modeled by utility function of appliances operation. Control is optimized on a sliding window and targets a well known finite control duration.

In [62] [72] and [87], aggregations of TCLs are represented as stochastic batteries. In [48], TCLs are controlled in real-time in a way that makes sure that the aggregate consumption of all active TCLs will always stay below a certain threshold. In this work, a deterministic rule based control is applied depending on an ordering of loads and predefined temperature bounds. Similarly, authors in [43] propose an interruptible load program scheme. A fuzzy logic based controller with a closed control loop is introduced to resolve prediction errors with the objective of minimizing operation costs. Based on previous work presented in [93] for load elasticity, [92] show the potential value of enabling power flexibility in appliances and proposes a distributed fuzzy logic based control with controllers deployed on appliance level that can react to congestion signals and try to reduce their power consumption to a desired set-point. In [1], the authors focus on the business model for aggregators to sell flexibility of an aggregation of users in ancillary services market while having some profit. In this case, users are clustered in groups that each have a certain consumption profile used to value the flexibility offered by these groups. Authors in [64] propose a distributed control schemes capable of providing ancillary services to the grid while utility of a homogeneous controlled appliances is optimized over an infinite horizon.

Fast DR

Fast DR schemes aim at providing frequency control to the power system (i.e., primary regulation) in the order of seconds. Most proposals suppose control approaches using a randomized policy to react to system frequency sensed locally. Authors in [14] propose to use randomization. Their proposal is motivated by the need to solve synchronization problems depicted when using the method described in [88]. Actually, when loads are used to provide primary regulation, a deterministic relationship between temperature limits of TCLs and frequency ([88]) used for setting frequency droop control can cause high swings of frequency when TCLs become synchronized. Authors in [75] propose an internal control model with two-way communication between the controller and TCLs . The controller broadcast a set-point temperature deviation to TCLs. The authors propose a decomposition of the reference signal into a number of signals to comply with resolution constraints for set-point temperature deviation imposed by thermostat hardware. They show that using proposed decomposition a trade-off is required between tracking performance and comfort. In [81], authors study the effect of DR on the performance of power systems based on their proportion in the system and communication delay. They show that without communication delay the DR considered approaches an accurate actuator that will increase system dynamic performance compared to traditional regulation resources. When delay is introduced, the actuator becomes less accurate. Finally, authors analyze required delay for a given penetration proportion of DR that will make the system under-perform compared to the traditional model without DR. This analysis is built on a simple model for DR response which gives insights into the effect of delay but do not discuss a possible elaborate scheme for load control.

Authors in [76] show the diverse effect that fast resources can have on system stability. They suppose the deployment of ideal storage system with unlimited capacity and instantaneous response. They show that with high proportions of batteries the dynamic performance of the system decreases. This is observed when fixing the response behavior to frequency deviation (i.e., droop parameters). Indeed, with high penetration of batteries, the system requires more time to stabilize frequency following a frequency disturbance. However, no discussion is made on the accuracy of the response of these batteries. Previously mentioned papers focus on grid modeling with no thorough discussion on appliances models.

4.2.2 Privacy

Direct load control is more likely to raise privacy issues compared to pricing-based DR. Indeed, this is mostly caused by its intrusive nature. To improve the acceptance of direct load control mechanisms privacy needs to be taken into account in addition to control performance and its capability of meeting system objectives. Indeed, depending on the control architecture and the distribution of intelligence, information flows may differ. Thus centralized approaches that optimize based on fine grained users information are usually the most privacy breaching approaches. Privacy issues can be mitigated when homes are equipped with a local intelligence capable of taking local decisions like in [35]. This limits or even prevents the need for sending sensitive users' information to aggregators.

In [25], authors propose to control appliances at a home to minimize the bill. However, the main focus is drawn on battery charging strategies that can hide consumption patterns of appliances. This makes it difficult for utility companies to extract private users information from their energy consumption. However, this proposal does not focus on the effect of control on users quality of experience. In addition, a simple appliance model is used where appliances are characterized by required energy consumption over optimization period.

4.2.3 Fairness

In most direct load control approaches, fairness is neglected while the main focus is based on performance. Fairness issues are addressed in some proposals like [51]. Indeed, the author propose real-time direct control schemes that allow to reduce the consumption of flexible devices below a certain desired threshold. Flexible devices are appliances whose consumption do not directly depend on user's activities at home. They introduce fairness by implementing a round robin control strategy.

5 Control of heterogeneous loads

While a significant number of studies has targeted residential users, very few focused on control schemes where different types of appliances are modeled. Indeed, most DR mechanisms targeting residential consumers have focused on specific devices (e.g., pools, heating, air-conditioning).

Among proposals that target heterogeneous resources, authors in [71] propose a two-stage approach: day-ahead pricing (so users can have an idea about the prices and use); real-time pricing (to adjust prices to real consumption). Authors model individual appliances to derive control decisions while also modeling distribution network constraints. No fairness mechanisms are implemented in this proposal. In addition, all users' preference information and constraints are sent to the aggregator. Authors in [97] propose a real-time control scheme that rewards users for accepting to control of their appliances. Users should provide all appliances related information to assess their importance for users and the flexibility they may offer depending on users defined constraints. The proposal of [89] is the most close to our modeling approach. However, in their approach, fairness is not addressed while privacy is introduced thanks to the distribution of intelligence. Indeed, based on their proposal, the aggregator optimizes the control signal based on appliances aggregate consumption information.

To reach the full potential of DR, heterogeneous demand modeling is of great importance and so is the control mechanism and architecture. Indeed, to reach out to residential consumers, there's a need to have a good understanding of the trade-offs underling the proposed services. In this thesis, we propose and analyze DLC mechanisms targeting heterogeneous loads taking into account (1) the impact on users and their satisfaction, (2) fairness, (3) privacy measures and (4) complexity in terms of the control algorithm and, information and communication requirements.

6 Summary and thesis motivation

DR is advocated as a mechanism that helps to optimize power grid operations and maintain generation-demand balance. It is a resource that can be used to substitute the need of generation capacity and to support the deployment of renewable energy sources. For DR services to reach their full potential, they should be able to deal with heterogeneity. This allows to jointly optimize demand generated from appliances providing different types of flexibility. While this is natural, a comprehensive flexible framework capable of managing heterogeneous appliances along with efficient control algorithms are still to be developed.

As for the possible control techniques, it has been shown that pricing-based DR has good potential in shaping demand. However, it does not provide strict guaranties to the system. In this thesis, we choose to focus on DR solutions based on direct load control approach for residential consumers. Specifically, we consider approaches that have slow and real-time decision timescale.

To encourage users to participate to such control, security concerns need to be addressed. Indeed, users usually worry about the use of their data and its impact on their privacy. So, limiting information disclosure or providing anonymity can increase the attractiveness of DR and hence the acceptance of such schemes. In addition, DR schemes need to take into account scalability as well as robustness and resilience constraints to provide an assured service. This encompasses communication delays and possible loss of information.

To reach previous requirements for DR service and fairness considerations, in the next chapters, we propose and analyze different ways of distributing intelligence among functional groups participating to the DR service. We evaluate proposed schemes by comparing the impact of resulting control strategies on users.

Chapter 3

Framework and utility functions

This chapter is dedicated to introducing a first, generic, system model on which we will build in subsequent chapters. Indeed, we will focus on presenting the general architecture we use to design direct load control DR schemes targeting residential users. This architecture will describe the main entities involved in the control and their possible interactions. Given any specific architecture, DR schemes ultimately controls demand at its finest granularity at the level of appliances. Control objective will be to lower the consumption of controlled appliance below a desired capacity limit over a targeted time period. In order to allow heterogeneous appliances to take part of the DR schemes, we propose a generic framework for describing appliances flexibility and usage through a taxonomy. Based on the taxonomy, we propose an approach to define the utility of operating an appliance as a function of monitored variables at homes. These functions are proposed to drive control decisions and assess their impact on user's perceived quality of experience. Based on the controlled appliances, decisions can be related to power delivery (e.g., turning ON or OFF an appliance) and/or the amount of power delivered (i.e. intensity).

This chapter is organized as follows: in Section 1, we present functional groups that will make up the architectural framework. We also introduce the families of DR solutions we study. We present our appliances taxonomy and describe how it is related to existing work on the subject in section 2. Based on the proposed taxonomy, we present a methodology to derive utility functions of appliances in section 3. We also give an example of an ideal planner capable of producing control policies that make use of these utility functions to optimize the impact of control on users' quality of service. Section 4 present examples of appliances that are modeled using the proposed methodology. These representative examples will be used in subsequent chapters to conduct numerical analysis on proposed DR solutions.

1 Framework

In this section, we present the architecture of the system we consider. Then, we introduce families of DR schemes that can be defined based on this architecture.

1.1 General architecture

Since the power distribution network is radial, we propose to model decisions with a similar structure. This allows to take into account possible physical constraints at different aggregation points of the network. Thus, we consider a hierarchical system of decision entities (made of N levels) that can be represented by a tree (see Figure 3.1). Each decision entity (say at a certain level of the hierarchy n) is a functional group responsible for a number of other decision entities that will make up the lower level (i.e., Level $n-1$). A decision entity can be seen as an aggregation point that can gather information and provide strict guaranties through direct control of the level(s) below. The decision entity on the top of the hierarchy (the N^{th} level) is responsible for deciding the global control target. Aggregation entities decide the target for the next levels; this target is usually dictated by the parent decision entity. Homes represent the lowest level of the hierarchy: only appliances stand below.

We argue that most of the challenges and opportunities offered by this multi-level architecture comes from the existence of at least two levels of decision. Thus, for the sake of simplicity, we suppose a two-level architecture in the rest of the dissertation. This will allow us through a simplified model to capture essential problems that arise when different levels of decisions are considered. We discuss the multi-level architecture for control schemes for which it's relevant.

Considered levels in this simplified model correspond to two functional groups: one at the Distribution System Operator (DSO) or aggregator side and one at the home side. We call them aggregator decision entity (DE_a) and home decision entity (DE_h) respectively.

For a given home, DE_h is aware of the state of the variables monitored at user premises and receives control decisions from DE_a. When relevant to a specific control case, DE_h can transmit all or part of the data he gathered to DE_a. Figure 3.2 illustrates the architecture considered.

1.2 DR solutions: centralized, partially distributed and distributed

Based on the general architecture presented previously, we define three families of DR solutions that we will address in the next chapters: centralized, partially distributed and distributed solutions. These families are introduced to propose DR solutions having different communication requirements. So to define the three families, we need to introduce communication specificities in our system.

As presented previously, we consider a system made of decision entities. Communication between these entities can be of two types: information (i.e., data) or orders (i.e., control signals). A decision entity has intelligence if it is capable of producing orders (control decisions).

In addition, communication can be *vertical* or *horizontal*. Communication is vertical when it involves entities at different hierarchical levels (see Figure 3.1). Such communication occurs when an entity at a certain level sends control signals to entities at lower levels, or when an entity sends data to another entity at a higher level. Vertical communication of control signals is required for all proposed DR approaches. This is due to global control targets being known only by the decision entity at the top of the hierarchy. Thus, they need to be propagated via control signals to entities at lower levels. Communication is horizontal when entities are on the same hierarchical level. It typically occurs when entities seek a certain consensus. This is achieved by exchanging data or contracts (e.g., transfer of capacity allocation).

Based on vertical communication of control signals, studied DR mechanisms can be classified in two families.

- In the first family, an aggregator decision entity directly controls the appliances (which represent the leafs of the hierarchy). To do so, it may either have full static or dynamic information on controlled appliances and relevant homes' variables. We refer to control schemes belonging to this family by centralized solutions.

- In the second family, an aggregator decision entity takes control decisions at the granularity of the decision entities in the level below. Decisions can be based on static information or dynamic aggregated information. So, controlling entity cannot have direct knowledge on the values of

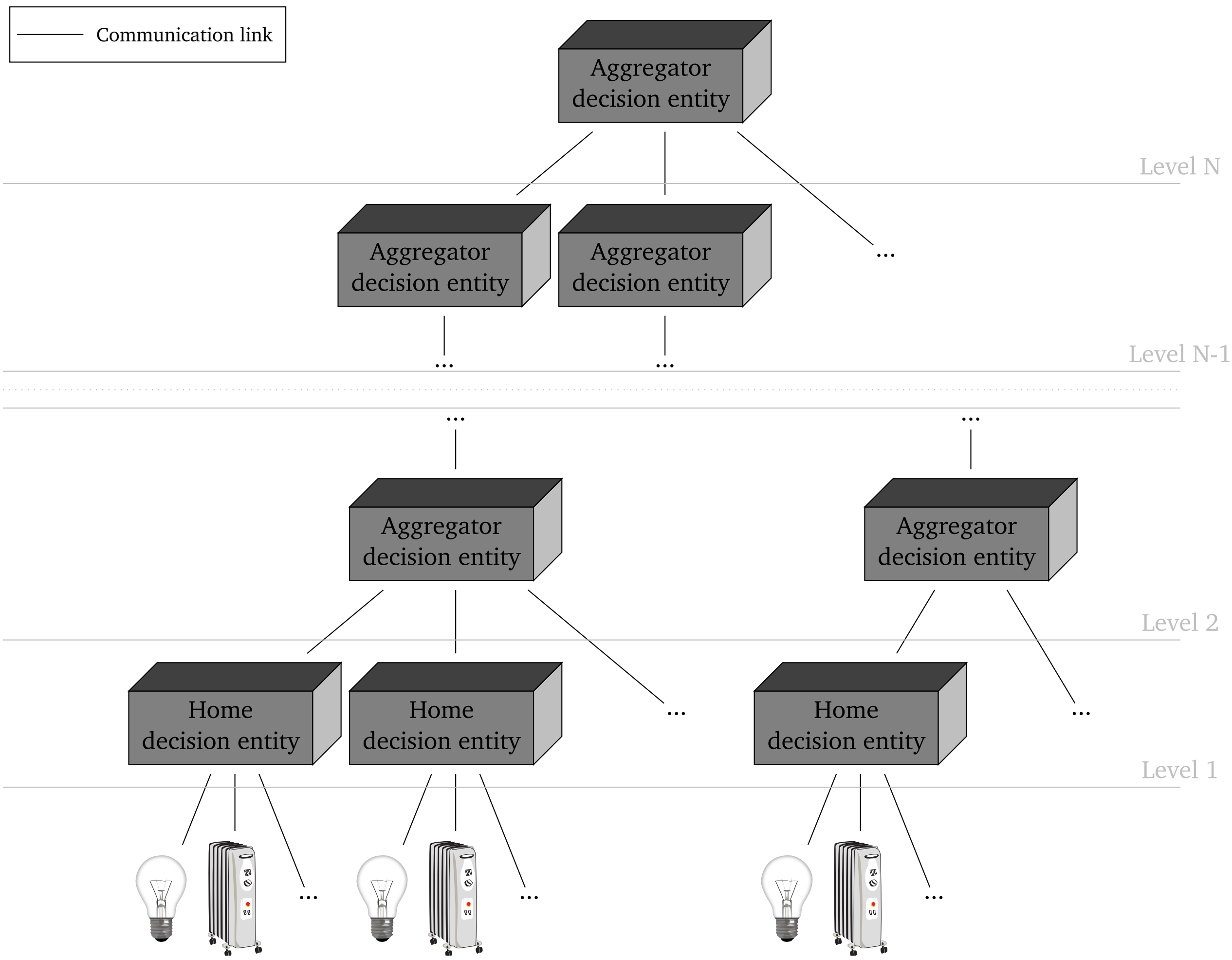


Figure 3.1 – Generic architectural framework

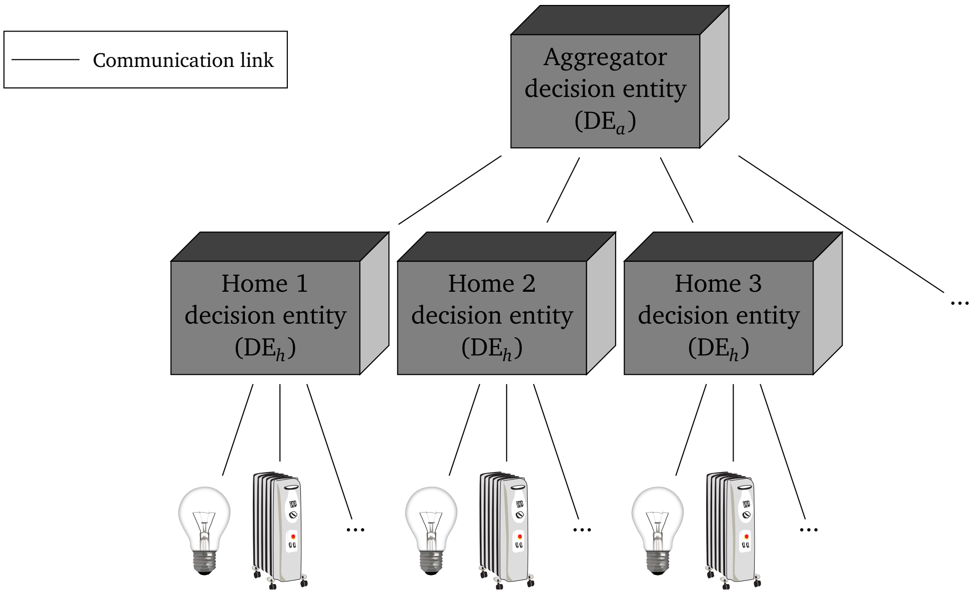


Figure 3.2 – Simplified architectural framework

individual appliances or the state of monitored variables at homes. Control schemes associated to this family are called partially distributed solutions.

As for horizontal communication, we refer to associated control schemes by distributed. This is due to the fact that no communication is required with upper levels.

As stated in the introduction of this chapter, we focus on time periods where it is predicted that global demand will exceed a desired capacity if no DR mechanism is deployed. Targeted capacity can be directly related to power grid infrastructure and operation constraints such as the lack of generation capacity or a high stress on system infrastructure (e.g., due to weather conditions like hot summer or cold winter). It can also be related to cost reduction objectives (e.g., reduce purchase on energy markets). We therefore target the proposal of DR solutions that aim lowering consumption below a desired capacity limit over a period of time. This period can be limited and well-known (i.e., finite horizon scenario) or it can be unlimited in which case the DR service can run at all times (i.e., infinite period).

Considering the two-level architecture from Figure 3.2, proposed centralized solutions consist of a DE_a controlling all the appliances on the network to satisfy the global capacity goal, and having full fine-grained information. In this case, DE_h at home will only relay control signals from DE_a to home's controlled appliances. DR mechanisms of this family allow us to reach the best possible solution due to the full visibility. However, fine-grained information disclosure may raise privacy concerns which can be addressed by the two other families of solutions through distribution of intelligence. For partially distributed solutions, DE_a takes decisions at the level of homes. So, it allocates to each home consumption limits (which could be different for different homes) based on the maximum available capacity for the considered time period. Based on the allocated maximum power, each DE_h takes decisions with respect to demands generated by each appliance the best way it can based on user's options. For this family, two possible scenarios are considered namely, one-way communication from DE_a to each DE_h or two-way communication between DE_a and each DE_h. Schemes based on one-way communication allow to keep private information disclosure to a minimum. However, if real needs cannot be known or estimated at the moment of decision, such schemes may result in control inefficiency. A two-way communication can potentially address visibility issues by giving some information on needs without fully breaching users privacy. For distributed schemes, DE_h at a home should communicate with other home decision entities to jointly optimize their consumption. These schemes allow us to have a system that is resilient to single point of failure.

In order to quantify energy efficiency and users satisfaction, we define for each appliance a utility function. The utility may depend on the monitored variables (like presence or temperature), on exogenous variables (like outdoor temperature) and of course on users' personal preferences. The DR models we introduce in the following chapters target maximizing the total utility under different types of system constraints and taking into account fairness considerations. We do not directly focus on the revenues of the players and try to adopt the point of view of an ideal planner. Nevertheless, one can expect that if well performing pricing models are defined, reaching maximum users' utility leads to maximum gains.

2 Appliances taxonomy

The impact of scheduling decisions on user's quality of experience depends on the nature of appliances being controlled. Thus, we propose a taxonomy that is general enough to describe any appliance's characteristics. We will rely on this taxonomy to capture an appliance's flexibility. Indeed, it will provide the rationale for the utility functions we propose afterward. These utility functions are used to design the optimization problems proposed in the next chapters.

2.1 Overview

In most related efforts, taxonomies proposed for classifying appliances depend on the DR service provided. They are mostly defined based on a desired demand feature the planner targets to exploit. So, a typical classification first separates appliances into:

- controlled (or controllable) appliances that are targeted by the DR service.
- uncontrolled appliances that should run freely or cannot be controlled.

Then, controllable appliances are further divided depending on the flexibility they can offer[1]. Demands generated by appliances can have features of being:

- **Shiftable/adjustable**. This means that operation can be shifted in time. This feature is typically exploited by users (or an energy box at their premises) when a pricing-based DR scheme is supposed to minimize their bills (see for example [6]). In [2], appliances with shifting capabilities are further separated into ones that have a known cycle of fixed duration (e.g., a washing machine) referred to by shiftable; others that can have a flexible cycle duration referred to by adjustable (e.g., a space heater).
- **Tolerable to certain interruption time period**. This feature is mainly considered in direct load control schemes to take consumption reduction decisions (e.g., see for example [54]).
- **Power flexible**. This feature describes the capability of changing the power rate consumed by the appliance from its nominal value (see [95] for examples). This is exploited by control to change the consumption level while avoiding complete interruption (e.g., [94]).

[1] In our analysis, we will consider a limited number of controllable appliances. However, the choice is not related to a specific flexibility provided by all appliances.

In these previously presented efforts, control is limited to the specific flexibility described by the classification. Each targeted flexibility can only be provided by certain appliances. These appliances are usually chosen so that control is seamless to end users (i.e., intrusiveness is minimized). However, in some cases, users may be willing to sacrifice part of their comfort in order to make sure certain important appliances get scheduled. So, the question asked when designing the DR schemes evolves from what should be cut or shifted to what should be scheduled. A simple example is heating during very cold winter when power is scarce. If heaters are the only controlled loads the outcome of the control might result in low temperatures of controlled homes compared to the case where users are asked to lower other demands in order to provide more capacity for heating. Hence, there's a need for a generic model that can describe the impact of power delivery for any generic appliance on users' perceived quality of experience.

Requirements needed to build a generic framework are partially addressed in several efforts. These requirements can be described by appliances usage, priority and operation features.

A classification centered on usage is provided by [55]. The authors propose four types of appliances. The first type is made of appliances that control the temperature of home's environment. The second type includes appliances that need to complete their tasks before a deadline. The third type includes appliances that have instantaneous effect on users (e.g., lighting). Finally, the fourth type include appliances that are not critical and for which users are concerned about their energy consumption (e.g., entertainment). This classification allows to derive main demand features from the first three types. However, priority is introduced as a separate type even though it is natural to suppose it between any appliances of the first three types. Fairness considerations and operation constraints are not discussed in this work.

In [50], priority is addressed by classifying appliances based on their criticality. Authors define three levels of criticality namely, "Critical", "Basic" and "Others". The first category comprises demands that have to be accommodated under any circumstances (e.g., home medical equipment). Basic demands are generated to accommodate important needs that will significantly affect users' well-being. (e.g., a certain level of lighting, space heating during especially cold weather). Demands that fall into the "Others" category are those that can be abandoned without significantly affecting home residents' well being (e.g. entertainment). This classification is centered on the objective importance of the appliances. However, consumers' class and preferences also need to be addressed as discussed in [98].

The technology of the appliance also affects the type of control that can applied to it regardless of its usage characteristics. Technology allows, for instance, to describe "power flexibility" discussed earlier. So, it is important to separate appliance's usage purpose from its technology. Several classifications are proposed based on power signature of appliances. Efforts in Non-Intrusive Load Monitoring (NILM) techniques (see [53, 57, 102]) show that, by measuring power consumption, it is possible to know the technology of the appliance being used (e.g., inductive, resistive, power electronic) along with other load characteristics. This classification allows us to derive operation constraints for appliances serving the same purpose (i.e., usage) but being based on different technologies.

2.2 Proposed taxonomy

For the purpose of modeling control mechanisms, we aim at separating the impact appliances have on users' perceived quality of experience from restrictions on control decisions. We therefore propose to characterize appliances based on the following three attributes, which stem from the previous classification proposals we have just seen.

Attribute 1 - Usage

This attribute will express expectations of users when operating the appliances. The value of this attribute is assigned by the user in case of ambiguity depending on the appliance's intended use. Inspired by [55], we propose grouping appliances based on 3 general types of usage:

- *Interactive demands* are those that are directly triggered by end users and, when possible, should be served immediately (similar to type 3 in [55]). Delaying the service of these demands is meaningless. Usage of appliances falling in this category usually depends directly on the presence of inhabitants. Lighting and entertainment (e.g., TV) are typical examples of interactive loads.

- *Background loads* are those for which users have a preference on the value of certain home variables affected by their power consumption but not directly on the power profile. They are usually used for controlling the value of variables at home like the temperature in a room, in the water-heater or inside the fridge (i.e., type 1 in [55]). They can also be used for controlling variables that are appliance specific (e.g., cleaning hours of pools, level of charge of batteries). These variables dynamics (i.e., how they evolve in time) depend on provided power and on exogenous variables (e.g. related with the weather). Demands generated by background loads are usually shiftable, which means that shifting (up to a certain limit) the time at which the expected energy is provided do not significantly affect the quality of experience and may be seamless for users. Indeed, the value of feeding such loads depends on the drift between the actual value of the variables being controlled and their desired value. It may also depend on user's presence at home (e.g., space heaters).

- *Program-based loads* are those that have well defined operation cycles that can be programmed. This usage is similar to type 2 in [55]. The value of feeding these appliances is related with the capacity of completing a program. They are usually shiftable. Indeed, an operation cycle can be scheduled between its earliest possible start time and its latest start time that allows it to complete before the expected deadline. Some appliances of this category may be stopped before the end of the cycle and still provide some utility. However, for some others, such an event may generate damage. We get back to this when we present operation characteristics (i.e., Attribute 3). A typical example of appliances belonging to this category is washing machines. Another example is ovens (which require each a certain operation duration to bake a meal).

To illustrate cases where a user may assign an appliance to different categories, we take Electric vehicle and television as examples. For an electric vehicle that can be charged and discharge (i.e., used as a battery), it can be seen as a background load if the user has a preferred level of charge. In this case, level of charge can be directly related to the average driving distance for the user so she can be able to drive the vehicle at any time (without having to specify the time of use of the vehicle). However, it can also be seen as a program-based load if the user sets a specific level of charge to be completed by a certain time deadline. In this case, the level of charge can be related, for example, to the distance between user's home and her workplace while the deadline corresponds to the latest time at which the user needs to leave his home. The same can also be applied for television. If the user demand targets a specific TV show that is aired at a specific time, Television can be seen as an interactive demand. However if user is willing to allow some time flexibility by specifying only the duration of the show (in case of TV on demand), television corresponds to a program based load.

Notice that the categorization based on usage allows us to extend related work description of the shifting flexibility. Indeed, usage features allows us to express the constraints on shifting. For example, shifting can be decided while taking into account the desire to maintain preferred values of certain variables in the case of background loads. It can also be determined in view of completing a task before a deadline in the case of program-based loads.

Attribute 2 - Criticality and preferences
As exposed in [50], some appliances may be more critical that others: for instance, home medical equipment have to operate when needed and are more critical than appliances used for entertainment. Moreover, this consideration is subjective and that is why users should be able to define themselves the criticality of their appliances.Furthermore, users may have targets corresponding to different criticality levels for a given appliance. So, control policies should enforce fulfilling targets with a certain criticality before trying to reach targets of lower criticality. Lighting and heating can illustrate such a scenario. Actually, it can be preferable to guaranty a certain temperature and a minimal level of lighting before augmenting the temperature to its preferred value or allowing all lighting system at home to function. Thus, in our proposal, we suppose different levels of criticality. Each appliances being controlled can fulfill one or more needs corresponding to the different levels. To capture the effect of having multiple levels for criticality per appliance, we will consider a simplified model based on two levels that we name *vital* and *comfort*. The first one expresses high priority targets of high impact on users' wellbeing (e.g., a certain level of indoor lighting, space heating during cold weather) and the second one expresses less essential preferences (e.g., warmer temperature, exterior lights). In addition to criticality, users may have different priority preferences of appliances serving the same criticality group.

Attribute 3 - Electrical and operation characteristics
This category allows us to distinguish appliances according to how they consume power in time. Constraints can be imposed by the specific technology of the appliance (e.g., resistive, inductive, storage). For a given appliances, we differentiate two types of flexibility: time related and power related.

Power flexibility describes whether it's possible to change/reduce the power rate of an appliance while allowing it to operate without hurting its lifespan. For resistive loads, reducing power will affect the quality of experience but not avoid functioning . In many cases resistive loads are elastic and therefore reducing allocated power will just delay the provision of the service (for example, a water-heater will take longer time to reach the desired water temperature). Inductive loads are usually less flexible. For example, a fridge compressor requires a certain power rate to function properly and to maintain its lifespan. Storage is usually power flexible. However, extensive cycling may hurt its capacity and efficiency.
Another case of flexibility is when one considers multiple instances of one given type of appliances as one big "meta appliance". An example of such simplification can be proposed for lighting. So providing a certain power values for lighting will enable to light up a number of bulbs which will provide different illumination levels.

Time flexibility of power allocation will describe a minimal duration of operation required. It can be defined to prevent harming the appliance or reducing its lifetime. Actually, certain loads like some inductive motors, impose a minimal duration of operation. It can also express required operation duration for non-preemptive demands that cannot be interrupted before completion of the task (e.g., washing machine).

3 System model

As mentioned in Section 1.2, our optimization problems rely on the definition of utility functions, which are deduced from the characteristics of considered appliances (presented in Section 2.2). Proposed methodology extends utility functions defined in [55] by including notions of home-related preferences and fairness. In this section, we start by presenting an approach for building utility functions. Then, we introduce the optimization problem we model with these functions. Notation used subsequently is summarized in Table 3.1.

3.1 Utility functions and constraints

We can take advantage of the taxonomy introduced in Section 2.2 to describe appliances we want to control and hence derive a set of utility functions and adequate control decisions for each controlled appliance. This is achieved in three steps :

1. Derive for each appliance its individual utility functions when considered alone (based on Attribute 1)
2. Introduce priority between multiple controlled appliances (based on Attribute 2)
3. Define constraints on control decisions for each appliance (based on Attribute 3)

We now detail each of the aforementioned steps.

3.1.1 Individual utility functions

When controlling appliances, power allocation needs to be decided in view of how users are impacted. To guide decisions, we want to design a generic framework independent of the nature of appliances being controlled and flexible enough to model any appliance. We also target to make it automatic while allowing users to manually set their subjective preferences.

To reach these goals, we define an individual utility function $f_{u_l ht}^{a}$ for each appliance a in home h at time t for criticality level u_l (discussed in Section 3.1.2). This function rates the instantaneous quality of experience as perceived by the user, based on the state of relevant home variables. The value of this function is defined between 0 (i.e. no perceived utility) and 1 (i.e., maximum perceived utility).

Since, when operated, each appliance impacts the user differently, the variables of the utility function need to be different. Thus, we consider for each appliance what we call *controllable variables*. These variables represent measurable quantities affected by the appliance's operation. A controllable variable can be defined by power consumption or it can be an exogenous variable affected by the appliance's operation (e.g., temperature). These controllable variables are considered to allow users to express their corresponding desired values (e.g., preferred temperature set-point) for which the perceived utility is maximal at the time instant (and at most equal to 1). Indeed, users may choose a maximal value for utility lower than 1 to express a lower interest in a specific time slot compared to others (e.g., powering on all lights is less interesting during the day than at night). Hence, the individual utility function is defined as follows:

Definition 2. *Let M_a be the number of controllable variables relevant to appliance a. We call individual utility function of appliance a at time t in home h for criticality level u_l , a function $f_{u_l ht}^{a} : \mathbb{R}^{M_a} \rightarrow [0, 1]$.*

Controllable variables relevant to an appliance can be determined based on the appliance's usage (attribute 1):

- For interactive loads, utility can be expressed as a direct function of power. This is due to the user's ability to instantaneously perceive any change in power allocated to the appliance.
- For background loads, utility can be expressed as a function of exogenous variables. These variables take values that depend on the power assigned to the appliance and other external parameters. A simple example is heating. Providing power to heaters will target to increase the value of indoor temperature (i.e., the controllable variable in this case). However, indoor temperature is also affected by the insulation of the home and exterior temperature.

- For program based loads, utility is related with the completion of a program. In this case, a possible controllable variable is power (constrained by program completion) or energy consumed by the appliance.

In Section 4, we instantiate a case study that will provide concrete examples of how utility functions can be defined based on the usage attribute.

So far, we only discussed utility function definition for a single appliance. However, when considering a collection of appliances in a single home or in multiple homes, control decisions need to deal with concurrent demand from several appliances. Hence, priority needs to be introduced in the utility function model.

3.1.2 Managing priority

At this point, appliances have the same scaling of utility functions. However, some appliances are more important than others. Hence, we need to introduce priority and criticality (Attribute 2 of the taxonomy defined in Section 2). They can be considered between homes, between appliances in the same home and between different operation targets of appliances (e.g, the importance of heating can be different depending on the temperature being in certain ranges). Such prioritization is important when available power cannot fulfill all appliances demand. We build our system such that supporting priority and criticality in the model is automatic once user's subjective priorities are expressed.

We recall that we use criticality to define strict differentiation between appliances operation importance. Priority is used to express ordering preference for appliances serving the same criticality level. This yields in two different requirements that need to be addressed in the utility function model: utility functions should not be comparable when they express needs of different criticality and they should be compared to arbitrate decision based on priority. Hence, next, we will treat the two concepts separately.

We now elaborate on how we evolve the utility function model from individual appliances utility to evaluating the impact when choosing among a set of appliances.

Priority

To define priority between appliances having the same criticality level u_l, we first need to take into account the scope of the control. Concretely, priority needs to be defined between appliances in the same house and between houses. This is achieved with proper scaling that allows us to define the actual instantaneous utility function ${U_{u_l}}_{ht}^{a}$ of an appliance a at home h at time t based the instantaneous individual utility function ${f_{u_l}}_{ht}^{a}$ (see Section 3.1.1). When proposing control schemes, control is decided in view of priority dictated by the actual instantaneous utility function of appliances. We now discuss the required scaling to move from individual utility functions to actual ones starting by scaling individual utilities of appliances inside a home then scaling obtained utilities accross homes.

For the control inside a home h, residents need to define for each appliance a a priority score ${\phi_{u_l}}^{a}(h)$. Priority scores are used to normalize individual utility functions in order to prioritize appliances inside the home. This also provides a flexibility for the utility of the same appliance to be of different priority for different homes.

To support homes with different quality of service requirements, a maximum utility $U_{u_l max}(h)$ is defined for each home h. We suppose that maximum utilities are set upon the establishment of contracts with the aggregator which use them to enforce priority between homes. So users having a high value $U_{u_l max}(h)$ will have a premium contract whereas homes for which maximum utility is low will have a more degraded quality of service.

Suppose that the control time period is slotted into t_M slots and A is the total number of controlled appliances. Based on individual appliance utility ${f_{u_l}}_{ht}^{a}$ and priority score of the appliance ${\phi_{u_l}}^{a}(h)$, the

System Parameters and Exogenous Variables	
τ	Time slot duration
H	Number of homes
A	Number of classes of appliances
M_a	Number of controllable variables affected by the operation of appliance a
$P_m^a(h)$	Minimum power for appliance a in home h
$P_M^a(h)$	Maximum power for appliance a in home h
$\pi_v^a(h,t)$	Vital preference coefficient for appliance a in home h at time t
$\pi_c^a(h,t)$	Comfort preference coefficient for appliance a in home h at time t
${\phi_{u_l}}^a(h)$	Priority score of appliance a at home h for criticality level u_l
$U_{u_l\,max}(h)$	Maximum utility of home h for criticality level u_l
${c_{u_l h}}^a$	Utility factor for appliance a at home h for criticality level u_l
t_M	DR period duration in time slots
$T_m(h)$	Minimum acceptable indoor temperature for home h
$T_P(h)$	Preferred indoor temperature for home h
$T_M(h)$	Maximum acceptable indoor temperature for home h
$T_0(h)$	Initial indoor temperature for home h
$F(h), G(h)$	Coefficients for temperature dynamics in home h
$T_e(t)$	Exterior temperature at time t
$T_s(h)$	Earliest start time of a washing machine at home h
$T_d(h)$	Deadline for completing washing machine operation at home h
$D(h)$	Duration of operation of the washing machine at home h
$C(t)$	Available power capacity at time slot t
Control Variables and Controlled Variables	
f_{ht}^a	Individual utility of appliance a in home h at time t
${f_{u_l ht}}^a$	Component of f_{ht}^a at criticality level u_l
U_{ht}^a	Actual utility of appliance a in home h at time t
${U_{v ht}}^a$	Vital component of U_{ht}^a
${U_{c ht}}^a$	Comfort component of U_{ht}^a
X_{ht}^a	Power consumed by appliance a in home h at time t
x_{ht}^a	$= 1$ when appliance a in home h at time t is active and $= 0$ otherwise
T_{ht}	Temperature of home h at time t

Table 3.1 – Table of notation

actual instantaneous utility function ${U_{u_l}}^a_{ht}$ of appliance a at home h at time t can be expressed by:

$${U_{u_l}}^a_{ht} = {c_{u_l}}^a_h \, {f_{u_l}}^a_{ht}$$

where ${c_{u_l}}^a_h = \frac{{\phi_{u_l}}^a(h) U_{u_l\,max}(h)}{max(\sum_{t=1}^{t_M} \sum_{a'=1}^{A} {\phi_{u_l}}^{a'}(h)\, {f_{u_l}}^{a'}_{ht})}$ is the utility factor for appliance a at home h. This utility factor allows us to impose a normalization that takes into account the maximum achievable utility for each house. Indeed, the denominator models possible constraints that prevents f^a_{ht} from reaching 1 all the time (e.g., no users at home, constraints that prevent one appliance from being operated when another is functioning). If all f^a_{ht} can reach 1, ${c_{u_l}}^a_h$ can be simplified to $\frac{{\phi_{u_l}}^a(h) U_{u_l\,max}(h)}{t_M \sum_{a'=1}^{A} {\phi_{u_l}}^{a'}(h)}$.

This enable to automatically distribute maximum utility over appliances based on the priority score. Indeed, when maximum individual utilities for appliances are reached at home h, total actual utility over the control time period reaches home's maximum utility $U_{u_l\,max}(h)$.

In the following chapters, we will be directly using actual utility functions ${U_{u_l}}^a_{ht}$ in the formulation of control algorithms targeting criticality levels u_l. We now discuss how we include criticality in the utility model.

Criticality

To deal with criticality, we suppose a multi-level decision model. Indeed, separating decisions is essential when dealing with appliances having different criticality levels. In fact, without separation, control policies can consider, for example, that lighting up a large number of light bulbs is equivalent in terms of utility to powering on a critical appliance (e.g. a dialysis machine). We further suppose that an appliance can serve different criticality purposes. So for each appliance, we define a tuple of actual utility values each corresponding to a level of criticality among criticality levels it serves. Suppose L levels of criticality. Suppose actual utility of appliance a at time t for home h is denoted by U^a_{ht}. We can write $U^a_{ht} = ({U_{u_1}}^a_{ht}, {U_{u_2}}^a_{ht}, .., {U_{u_L}}^a_{ht})$ where u_1, u_2, .., u_L express criticality levels in decreasing order.

When controlling a set of appliances, policies target to satisfy most important needs expressed by u_1 utility level before improving other utility levels. The same apply to other utility levels where each level has to be fully optimized before improving utility values of lower levels.

As stated in Section 2.2, we suppose a simplified two levels model of utility per appliance, named *vital* and *comfort*.

So, we can write utilities as vital/comfort pairs: $U^a_{ht} = ({U_v}^a_{ht}, {U_c}^a_{ht})$. Therefore, control policies target to satisfy comfort only if vital needs cannot be further covered for all appliances. Control decisions are based on the lexicographical order comparison of utility values:

For two utility values U^a_{ht} and U'^a_{ht}, we say

$$U^a_{ht} > U'^a_{ht} \text{ iff } {U_v}^a_{ht} > {U'_v}^a_{ht} \text{ or } ({U_v}^a_{ht} = {U'_v}^a_{ht} \text{ and } {U_c}^a_{ht} > {U'_c}^a_{ht}).$$

Utilities can be summed using element-wise addition.

3.1.3 Control type

The optimization problem we target to solve consists in deciding the amount of power assigned to each appliance (see Section 3.2). Decision quality is deduced based on the actual utility function defined following Sections 3.1.1 and 3.1.2.

Based on attribute 3 in Section 2.2, we can identify the type of control that can be applied. Decision can be discrete (i.e., ON-OFF) or continuous (i.e., power intensity). Additional time related constraints may be considered to model tolerance to interruptions (e.g., once the appliance is operated, it should remain ON for a certain duration). This limits possible utility improvement to admissible ranges of power.

For any appliance a at home h, power consumption constraints at a certain time instant t can be formulated using two parameters and two decision variables. The parameters defined in this case are a minimum power value $P_m^a(h)$ and a maximum power value $P_M^a(h)$ in Watts. Two specific cases can be noted. In case of appliances with continuous power only limited by a maximum rate (e.g., batteries), minimum power parameter $P_m^a(h)$ is equal to zero and maximum rate is denoted by $P_M^a(h)$. For appliances that only allows ON-OFF decisions, maximum power $P_M^a(h)$ is equal to minimum power $P_m^a(h)$.

Decision variables are expressed by power X_{ht}^a allocated to appliance a at home h at time t and binary ON-OFF decision variable x_{ht}^a (equal to 1 when appliance is ON and to zero when it is OFF).

This formulation can be expressed by equation (3.1).

$$P_m^a(h)x_{ht}^a \leq X_{ht}^a \leq P_M^a(h)x_{ht}^a. \tag{3.1}$$

Additional constraints on power allocations can be imposed. An example of such constraints is given in Section 4.3 for washing machines. In this example, we impose that once the appliance is operated it cannot be stopped before completing it's program.

3.2 Optimization problem

We now present an optimization problem based on utility functions and power constraints presented in Section 3.1. This problem can be solved by a home decision entity or an aggregator (see Figure 3.2). Utility functions are defined to guide control decisions when allocating power among a set of appliances. We can define an ideal planner by being a decision entity that controls appliances in a way that users' satisfaction is maximized. Suppose a targeted control period t_M, a set of H homes (H can be equal taken equal 1) and a set of A appliances. To limit decision times, we suppose time is slotted into intervals of equal duration τ. Since an ideal planner will try to optimize users satisfaction on the network, the objective function of such a planner can be formulated by equation (3.2).

$$\max_{X_{ht}^a, x_{ht}^a} \sum_{t=1}^{t_M} \sum_{h=1}^{H} \sum_{a=1}^{A} U_{ht}^a \quad . \tag{3.2}$$

As stated in Section 1.2, the control targets to lower the consumption below a certain desired capacity value during the targeted control period. This control target can be formulated by equation (3.3).

$$\sum_{h=1}^{H} \sum_{a=1}^{A} X_{ht}^a \leq C(t), \;\; \forall\, t. \tag{3.3}$$

Each variable X_{ht}^a represents power allocation decision for appliance a at home h for time slot t. In the following chapters, we will be proposing and resolving control problems that target to reach the performance of an ideal planner.

When multiple levels of utility are considered (e.g. vital and comfort), the problem needs to take into account the lexicographic ordering of utility functions. A complete formulation of the optimization problem is provided in Section 4.4 for a specific example of controlled appliances considering vital and comfort utilities.

4 Case study

The goal of this section is to show how our model can be instantiated for heterogeneous appliances (i.e. falling in different categories of our taxonomy) and hence to demonstrate its tractability and its genericity.

To do that, we need to make choices on the appliances that will be studied and the priority/criticality we give to each one of them. Since, it is not possible to study all possible choices for both cases, we restrict the analysis to:

- a presentative subset of appliances: we choose it by covering all types of usage described by our taxonomy (see Section 2) while keeping it simple by supposing one controllable variable for each appliance. We will specifically study lighting, heating systems and washing machines as examples of interactive, background and program-based loads respectively.

- a neutral case study where no priority preference is expressed between homes and appliances. So, we suppose a case study where all homes have the same maximum utility (i.e., no service differentiation). In addition, home's maximum utility is equally distributed over appliances (i.e., all appliances have the same maximum total actual utility).

- a two-level utility model for appliances where each level correspond to a different criticality level: one level corresponds to fulfilling vital needs and the second one corresponds to comfort. So, for each appliance a in home h at time t, we will define a vital utility function ${U_v}^a_{ht}$ and a comfort utility function ${U_c}^a_{ht}$.

In each section, we show the steps required to define actual utility functions for a given appliance. We model individual utility by piecewise linear functions because such functions allow us in a simple way to approximate the perceived value from an appliance's usage while enabling efficient resolution of the optimization problems. As discussed in Section 3.1.1, the maximal value of this function can be taken at most equal to 1. We define $\pi_v^a(h,t) \leq 1$ and $\pi_c^a(h,t) \leq 1$ that we call vital and comfort preference coefficients for appliance a respectively to illustrate cases in which a user may decide to express for the same appliance different interests in different time slots.

We now illustrate how to instantiate our framework for the relevance appliances we identified. We assign an index $a \in A$ for each appliance, namely $a = 1$ for lighting systems (presented in Section 4.1), $a = 2$ for heating systems (presented in Section 4.2) and $a = 3$ for washing machines (presented in Section 4.3). We remind that notation can be found in Table 3.1, page 35.

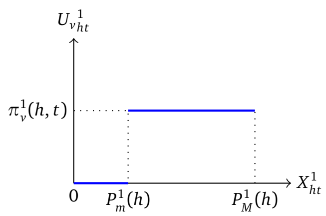
4.1 Lighting System

Description

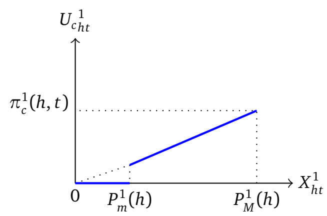
A lighting system is considered as an interactive load example (based on Attribute 1). Its criticality is related to the level of illumination required by users (Attribute 2) which is directly related to power consumption. Power drawn for lighting depends on the number of bulbs powered on and on the power they consume. So operation constraints are related to the power that can be consumed by bulbs to be lighted up (Attribute 3).

Operation constraints

Power allocation for lighting is either 0 (i.e., OFF) or a value between minimum power and maximum power denoted by $P_m^1(h)$ and $P_M^1(h)$ respectively. The minimum power can be chosen to be equal to minimal lighting requirements (e.g., to ensure users security at home). The maximum power is given by the sum of the power of all the lighting system at the home.



(a) Vital utility



(b) Comfort utility

Figure 3.3 – Utility of light power

Utility model

Individual utility model

We consider a linear individual utility function for lighting. To express a vital need in providing minimum power required by lighting system to be powered on, individual vital utility $f_{v\,ht}^{\,1}$ is considered such that its maximum is reached if minimum power $P_m^1(h)$ is provided. For comfort individual utility, we suppose that user's comfort is directly related to how many bulbs can be lighted up and this comfort is maximized when all bulbs can be powered on. So, for comfort, we consider a linear function that grows when power allocated increases from minimum power $P_m^1(h)$ to maximum power $P_M^1(h)$.

Additional power values in the interval $[P_m^1(h), P_M^1(h)]$ can be supposed, for example, to model maximum lighting power required for certain times of the day based natural sunlight. Here, we do not consider such intermediary values for power. However, we can imagine that these values can be determined based on illuminance requirements (see for example [41]). Indeed, illuminance desired can be converted to irradiance in Watts/m^2. Multiplying irradiance by the surface of the home (or the surface of a targeted area that needs to be illuminated) gives a value in Watts corresponding to needed power for optimal illumination of the area(s).

We suppose that maximal vital and comfort values of individual utility functions $f_{v\,ht}^{\,1}$ and $f_{c\,ht}^{\,1}$ are given by $\pi_v^1(h,t)$ and $\pi_c^1(h,t)$ which express vital and comfort preference coefficients respectively. These coefficients can be set by the user to change the utility perceived in different time slots (e.g., 1 can express her presence at home, 0 can represent having enough natural sunlight during certain time of the day).

Priority

To move from individual utility functions to actual ones, we need to define utility factors $c_{v\,h}^{\,1}$ and $c_{c\,h}^{\,1}$ (see Section 3.1.2).

In this example, since we aim at taking a neutral case study and to make it simple for this first appliance, we take $c_{v\,h}^{\,1} = 1$ and $c_{c\,h}^{\,1} = 1$ which gives $U_{v\,ht}^{\,1} = f_{v\,ht}^{\,1}$ and $U_{c\,ht}^{\,1} = f_{c\,ht}^{\,1}$. We define the scale of the other appliances based on the maximum total utility for lights. Suppose t_M denotes the control period in number of slots. This gives a maximum total utility for lights equal to $\sum_{t=1}^{t_M} \pi_v^1(h,t)$ for fulfilling vital needs and equal to $\sum_{t=1}^{t_M} \pi_c^1(h,t)$ for fulfilling comfort. Actual utility function for lighting is given by Figure 3.3.

MILP formulation

We now derive inequalities that are considered in the formulation of utility functions in a Mixed Integer Linear Problem (MILP) framework that can express the previously described individual utilities while scaling them depending on priority. We define variables for lighting that will be used to express its actual utility. Since, DR algorithms that control appliances will target utility maximization, functions for actual utility can be represented as upper bounds of the utility variables.

To formulate the utility functions in Figure 3.3, we define two types of utility variables for lights which are $U_{v_{ht}}^{1}$ and $U_{c_{ht}}^{1}$ that denote vital and comfort utility respectively.

Variables $U_{v_{ht}}^{1}$ can be expressed in terms of binary decision variables x_{ht}^{1} (= 1 for ON and 0 for OFF) following equation (3.4).

$$U_{v_{ht}}^{1} = \pi_{v}^{1}(h,t)x_{ht}^{1}, \ \forall\, t, \ \forall\, h. \tag{3.4}$$

The right hand side of equation (3.4) expresses the linear vital utility function in the interval $[P_{m}^{1}(h)$,$P_{M}^{1}(h)]$ (see Figure 3.3a).

Comfort utility variables $U_{c_{ht}}^{1}$ can be expressed by equations (3.5).

$$U_{c_{ht}}^{1} = \frac{\pi_{c}^{1}(h,t)}{P_{M}^{1}(h)} X_{ht}^{1}, \ \forall\, t, \ \forall\, h. \tag{3.5}$$

The right hand side of equation (3.4) expresses the linear comfort utility function in the interval $[P_{m}^{1}(h)$,$P_{M}^{1}(h)]$ (see Figure 3.3b).

4.2 Heater System

Description

A heater system is considered as a background load example that controls indoor temperature (based on Attribute 1). However, indoor temperature value evolves following a heat transfer equation. So other parameters that are uncontrollable directly by power like insulation, opened windows and exterior temperature affect its values. We suppose that these parameters can be predicted or collected by the entity responsible of controlling the heater. The criticality of heating depends on the indoor temperature required for users (Attribute 2). As for operation constraints (Attribute 3), we suppose that heating system can be made up of one (with adjustable power rate) or more heaters that can be turned on.

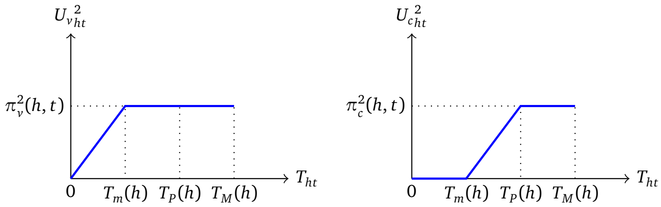
Operation constraints

An admissible power value is supposed to be either 0 (i.e., OFF) or a value between a minimum heating power $P_{m}^{2}(h)$ and a maximum power $P_{M}^{2}(h)$ for a given home h.

Utility model

Individual utility model

For a heater, we consider that vital needs are fulfilled when temperature is higher than a minimum tolerable temperature. So, vital utility $f_{v_{ht}}^{2}$ linearly grows until the minimum tolerable temperature $T_{m}(h)$ is reached. Its maximal value is represented by the preference coefficient $\pi_{v}^{2}(h,t)$. Comfort utility $f_{c_{ht}}^{2}$ linearly increases from zero at the minimum acceptable temperature $T_{m}(h)$ to the maximum value $\pi_{c}^{2}(h,t)$ at the preferred temperature $T_{P}(h)$. Coefficients $\pi_{v}^{2}(h,t)$ and $\pi_{c}^{2}(h,t)$ can represent preferences related to users' presence at home or to defining different weights for temperature values depending on the time of the day (e.g., daytime or night). They are defined in the interval [0,1].



(a) Vital utility (b) Comfort utility

Figure 3.4 – Utility of temperature

Temperature evolution dynamics can be modeled by the following first order differential equation when time is continuous:

$$\dot{T}(t) = fX + g(T_e(t) - T(t))$$

.

In this case, X represents power delivered by the heating system at time t. Exterior temperature is denoted by $T_e(t)$. Parameter f represents the inverse of thermal capacity and parameter g expresses the inverse of the multiplication of thermal capacity by thermal insulation. For a constant exterior temperature T_e, the solution of the differential equation is given by:
$T(t) = (T_0 - T_{eq})e^{-gt} + T_{eq}$ where $T_{eq} = T_e + \frac{fX}{g}$ denotes equilibrium temperature and T_0 is the initial indoor temperature.

So, we can see that without heating temperature decreases exponentially to reach equilibrium temperature (T_e in this case) when $t \to \infty$.

In addition, heating at time slot t will generate temperature increase that will affect subsequent time slots (due to inertia). For subsequent time slots, temperature degrees due to heating evolves according to:
$\hat{T}(t) = T(t) + (\hat{T}_0 - T_0)e^{-gt}$ where $\hat{T}(t)$ and $\hat{T}_0$ denote temperature at time $t \geq 0$ and at time $t = 0$ respectively when heating is activated at time $t = 0$.

We can see that temperature increase generated by heating will decay exponentially with time to reach zero when $t \to \infty$.

Since we are supposing a discrete case, dynamics can be expressed by the following equation:
$T_{ht} \approx T_{h(t-1)} + F(h)X^2_{ht} + G(h)(T_e(t) - T_{h(t-1)})$ where T_{ht} is the temperature in home h at time t and $T_e(t)$ is the exterior temperature at time t. X^2_{ht} is power delivered for heating in home h at time t. $F(h)$ and $G(h)$ are coefficients for heating in home h that can be derived from the continuous time dynamics by multiplying f and g by time slot duration τ respectively. Temperature at the beginning of the optimization time period is known and denoted by $T_0(h)$. To prevent excessive heating, interior temperature cannot exceed a maximum value $T_M(h)$.

Priority

Similarly to priority defined for lighting, we suppose that the heating system can reach the same maximum total actual utility as lights. So, we need to define the utility factor that allows to obtain actual heater utility from individual utility functions.

We suppose that both lights and heaters have the same possible total utility: $\sum_{t=1}^{t_M} \pi^1_v(h,t) = \sum_{t=1}^{t_M} \pi^2_v(h,t)$ for vital needs and equal to $\sum_{t=1}^{t_M} \pi^1_c(h,t) = \sum_{t=1}^{t_M} \pi^2_c(h,t)$ for comfort ones. In this case, we can also take $U^2_{vht} = f^2_{vht}$ and $U^2_{cht} = f^2_{cht}$. Actual utility functions for heating are given in Figure 3.4.

MILP formulation

Now, we define utility variables constraints that can be used in a MILP programming framework. We suppose that temperature variables $T_{ht}, \ \forall\, t, \forall\, h$ follow the previously presented discrete case dynamics.

For modeling heaters actual utility functions, vital and comfort utility variables are defined and denoted by $U^2_{v_{ht}}$ and $U^2_{c_{ht}}$. We now define upper bounds of these variables that represent functions in Figure 3.4.

Vital utility can be expressed by equation (3.6).

$$U^2_{v_{ht}} \leq \frac{\pi^2_v(h,t)}{T_m(h)} T_{ht}, \ \forall\, t, \forall\, h \tag{3.6a}$$

$$U^2_{v_{ht}} \leq \pi^2_v(h,t), \ \forall\, t, \forall\, h \tag{3.6b}$$

The right hand side of equation (3.6a) expresses the linear function defined in the interval $[0, T_m(h)]$ in Figure 3.4a whereas equation (3.6a) denotes the linear function defined in the interval $[T_m(h), T_M(h)]$. Indeed, if temperature T_{ht} is lower than $T_m(h)$, admissible values of utility are given by the area under the function $\frac{\pi^2_v(h,t)}{T_m(h)} T_{ht}$. Since we aim at maximizing utility, values considered for utility variable $U^2_{v_{ht}}$ will be the ones on the linear function. Temperature variables T_{ht} cannot be higher than $T_M(h), \ \forall\, t, \forall\, h$. This limit is imposed so homes are not used as virtual batteries in a way that hurts residents' well-being.

To model comfort utility, we need to deal with a function that is piecewise linear and non-convex. Artificial binary variables denoted by y_{ht} need to be introduced to model the non-convexity for temperature equal to $T_m(h)$. Variables y_{ht} take the value 1 when the temperature T_{ht} is greater than minimum acceptable temperature $T_m(h)$ and 0 otherwise. One simple possible way to model this function is through the Big M method. Indeed, the Big M method is a technique that allows to emulate conditional equalities. To give an idea about how it can be used, we suppose a large number M (representing infinity). We can define comfort utility variables by equation (3.7).

$$U^2_{c_{ht}} \leq \frac{\pi^2_c(h,t)}{T_P(h)-T_m(h)} (T_{ht} - T_m(h)) + M(1 - y_{ht}), \ \forall\, t, \forall\, h \tag{3.7a}$$

$$U^2_{c_{ht}} \leq \pi^2_c(h,t) y_{ht}, \ \forall\, t, \forall\, h \tag{3.7b}$$

When temperature T_{ht} is less than $T_m(h)$ ($y_{ht} = 0$), comfort utility has to be equal to zero. This is done using M to render inequality (3.7a) invalid ($U^2_{c_{ht}} \leq +\infty$). The larger the number M is taken the less efficient the formulation is. So, inequalities need to be optimized to produce a tight formulation. Working towards this goal, we can write an optimized formulation expressed by equation (3.8).

$$U^2_{c_{ht}} \leq \frac{\pi^2_c(h,t)}{T_P(h)-T_m(h)} (T_{ht} - T_m(h) y_{ht}), \ \forall\, t, \forall\, h \tag{3.8a}$$

$$U^2_{c_{ht}} \leq \pi^2_c(h,t) y_{ht}, \ \forall\, t, \forall\, h \tag{3.8b}$$

In Figure 3.4b, equations mapping to the function is as follows:

- The right hand side of equation (3.8a) expresses the linear function in the interval $[T_m(h), T_P(h)]$ (for which $y_{ht} = 1$).
- The right hand side of equation (3.8b) expresses the linear function in the interval $[T_P(h), T_M(h)]$ (for which $y_{ht} = 1$).
- For temperatures less than the minimum acceptable ($y_{ht} = 0$), both equations (3.8a) and (3.8b) will enforce a utility value of zero.

4.3 Washing machine

Description

A washing machine is a program based load for which a starting time and a deadline are set by the user (based on Attribute 1). Criticality is expressed in terms of being able to operate and consume certain amount of energy before the deadline (Attribute 2). It is modeled by a block of power spanning a certain time period. The block has to be scheduled in a time range defined by an earliest start time and a deadline. Once the block is scheduled it cannot be interrupted (Attribute 3).

Operation constraints

Power allocation for a washing machine at home h is either 0 (i.e., OFF) or $P_m^3(h) = P_M^3(h)$ (i.e., ON). We also impose additional precedence constraints. Based on these constraints, if the appliance is turned ON, it should remain ON till the end of its operation. Suppose $t_s(h)$ is the earliest start time and $t_d(h)$ is the deadline by which the operation should end. Suppose $D(h)$ is the operation period of the washing machine. ON-OFF power decisions are denoted by the binary variable x_{ht}^3 (i.e., $x_{ht}^3 = 0$ for powering OFF and $x_{ht}^3 = 1$ for powering ON). We impose that for time instants $t < t_s(h)$ or $t > t_d(h)$, we cannot operate the appliance ($x_{ht}^3 = 0$).

For time instants $t \in [t_s(h), t_d(h)]$, to enforce an uninterrupted operation, we should take into account three time intervals: the first $D(h)$ time slots starting from the earliest start time, the latest $D(h)-1$ time slots to reach the deadline and the time interval in between. This is done as follows:

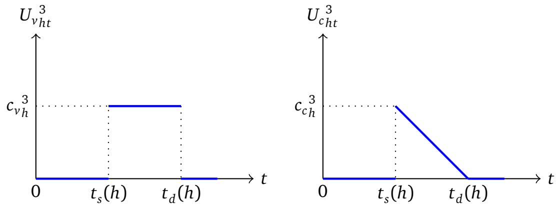
- In the first $D(h)$ time slots starting from the earliest start time (for $t < (t_s(h) + D(h)))$, once the appliance is operated, it should remain ON for the rest of this interval. This statement can be enforced by the inequality $x_{ht}^3 \geq x_{h(t-1)}^3 \; \forall t < (t_s(h) + D(h)$. Indeed, once x_{ht}^3 equals 1 at a given t, the variables corresponding to subsequent time slots are necessarily equal to 1.
- In the latest $D(h)-1$ slots to reach the deadline (for $t > (t_d(h) - D(h) + 1)$), if the appliances is not operated, it will not be scheduled. This statement can be enforced by the inequality $x_{ht}^3 \leq x_{h(t-1)}^3$. Indeed, if x_{ht}^3 at $t = t_d(h) - D(h) + 1$ is equal to zero, all power binary decision variables will be equal to zero $\forall t > (t_d(h) - D(h) + 1)$.
- For time slots $t \in [t_s(h) + D(h), t_d(h) - D(h) + 1]$, we need to make sure that the washing machine can be operated at any time for exactly $D(h)$ time slots. Precedence can be enforced by inequalities $x_{ht}^3 \geq x_{h(t-1)}^3 - x_{h(t-D(h))}^3 \; \forall t \in [t_s(h)+D(h), t_d(h)-D(h)+1]$. Duration constraint can be formulated by $\sum_{t=1}^{t_M} x_{ht}^3 \leq D(h)$. Indeed, inequality is used for the duration constraint because we allow situations in which the washing machine is not scheduled.

Utility function

Individual utility model

Individual utility functions are defined to express user preferences for operation and its time of schedule. So, as a vital need, we suppose a constant individual utility function $f_{v_{ht}}^3 = 1$ for $t \in [t_s(h), t_d(h)]$ when the appliance is operated. This will enforce the scheduling of the appliance between the earliest start time ($t_s(h)$) and the deadline ($t_d(h)$). In the proposed model, we take a comfort utility function $f_{c_{ht}}^3$ that reflects home's preference for the scheduling of the appliance. Indeed, we suppose a comfort utility function that expresses home's preference to start the washing machine as soon as possible . So its maximum value is reached when the appliance is scheduled at its earliest starting time ($f_{c_{ht}}^3 = 1$ for $t = t_s(h)$) and the lowest when the appliance finishes by the deadline ($f_{c_{ht}}^3 = 0$ for $t = t_d(h)$).

With the supposed example, individual utility functions are defined as functions of time. However, their values are constrained with the fact that the washing machine is operated. So, utility variables



(a) Vital utility (b) Comfort utility

Figure 3.5 – Utility of a washing machine

are expressed by the function value multiplied by binary decision variables x_{ht}^{3} (since power decision are restricted to ON-OFF in the example we take).

Priority

We can see that as opposed to lighting system and heating, individual utility of a washing machine is limited to the operation duration $D(h)$. When we define actual utility functions of the washing machine ($U_{v\,ht}^{3}$ for vital and $U_{c\,ht}^{3}$ for comfort) based on individual utility functions ($f_{v\,ht}^{3}$ and $f_{c\,ht}^{3}$ for vital and comfort respectively), we need to make sure that appliance's maximum utility values are equal to the other appliances' maximum values (neutral scenario). This requires a scaling of individual utility functions.

Indeed, based on individual utility functions, the maximum total vital utility is equal to operation duration $D(h)$ whereas the maximum total utility for comfort is less than $D(h)$. To adjust utility functions supposing a control period t_M, we define two utility factors $c_{v\,h}^{3}$ and $c_{c\,h}^{3}$.

To reach the same maximum vital utility as lighting and heating, we should define $c_{v\,h}^{3}$ such that $D(h)\, c_{v\,h}^{3} = \sum_{t=1}^{t_M} \pi_v^1(h,t) = \sum_{t=1}^{t_M} \pi_v^2(h,t)$. In subsequent chapters, numerical analysis are conducted supposing that preference coefficients for lighting and heating systems ($\pi_{v/c}^{a}(h,t)$) are set to 1 for both vital and comfort utilities. In this case, individual vital utility function need to be multiplied by:

$$c_{v\,h}^{3} = t_M / D(h).$$

For comfort utility adjustment, a multiplying factor needs to be computed such that the summation of the first $D(h)$ utility values (i.e., $\sum_{t=t_s(h)}^{t_s(h)+D(h)-1} c_{c\,h}^{3}\, max(f_{c\,ht}^{3})$) is equal to $\sum_{t=1}^{t_M} \pi_c^1(h,t) = \sum_{t=1}^{t_M} \pi_c^2(h,t)$. In the considered numerical analysis case ($\pi_c^a(h,t) = 1\ \forall h \forall t$), comfort utility factor is given by:

$$c_{c\,h}^{3} = \frac{t_M(t_d(h) - t_s(h))}{D(h)((t_d(h) - t_s(h)) - \frac{D(h)-1}{2})}.$$

Actual utility functions for a washing machine at home h is represented in Figure 3.5).

MILP formulation

To model utility functions in a MILP framework, we define vital and comfort utility variables $U_{v\,ht}^{3}$ and $U_{c\,ht}^{3}$ respectively.

To formulate the function in Figure 3.5a, Vital utility variables $U_{v\,ht}^{3}$ can be expressed by equation (3.9).

$$U_{v\,ht}^{3} = c_{v\,h}^{3} x_{ht}^{3}, \quad \forall\, t, \forall\, h \tag{3.9}$$

As stated, vital utility perceived depend on operation decision. Indeed, once the appliance is scheduled, maximum vital utility in reached for an starting time of the operation.

Based on Figure 3.5b, comfort utility variables $U_{c\,ht}^{3}$ can be defined by equation (3.10).

$$U_{c\,ht}^{3} = \frac{t_M\ (t_d(h) - t)}{D(h)((t_d(h) - t_s(h)) - \frac{D(h)-1}{2})} x_{ht}^{3} \forall t \in [t_s(h), t_d(h)]. \tag{3.10}$$

Equation (3.10) implies that maximum utility is obtained when the appliances is operated at its earliest start time. In addition, based on equations (3.9) and (3.10), no utility is perceived if the appliance is not scheduled.

4.4 Summary

In this section, we gave three examples of appliances that are modeled through the proposed generic framework. Considering a vital/comfort utility model, the problem solved by an ideal planner that controls all three appliances will be made of two steps.

In a first step, vital utilities are optimized by solving the MILP problem formulated by equations (3.11).

$$\max_{X_{ht}^a, x_{ht}^a} \sum_{t=1}^{t_M} \sum_{h=1}^{H} \sum_{a=1}^{A} U_{v\,ht}^{\ a} \tag{3.11a}$$

s.t.

$$\sum_{t=1}^{t_M} \sum_{h=1}^{H} \sum_{a=1}^{A} U_{v\,ht}^{\ a} \geq U_v \tag{3.11b}$$

$$\sum_{h=1}^{H} \sum_{a=1}^{A} X_{ht}^{a} \leq C(t), \ \ \forall\ t \tag{3.11c}$$

$$P_m^a(h) x_{ht}^a \leq X_{ht}^a \leq P_M^a(h) x_{ht}^a, \ \ \forall\ t,\ \forall\ h,\ \forall\ a \tag{3.11d}$$

$$U_{v\,ht}^{\ 1} = \pi_v^1(h,t) x_{ht}^1, \ \ \forall\ t,\ \forall\ h \tag{3.11e}$$

$$T_{ht} = T_{h(t-1)} + F(h) X_{ht}^2 + G(h)(T_e(t) - T_{h(t-1)}), \ \ \forall\ t,\ \forall\ h \tag{3.11f}$$

$$T_{ht} \leq T_M(h), \ \ \forall\ t,\ \forall\ h \tag{3.11g}$$

$$U_{v\,ht}^{\ 2} \leq \frac{\pi_v^2(h,t)}{T_m(h)} T_{ht}, \ \ \forall\ t,\ \forall\ h \tag{3.11h}$$

$$U_{v\,ht}^{\ 2} \leq \pi_v^2(h,t), \ \ \forall\ t,\ \forall\ h \tag{3.11i}$$

$$x_{ht}^3 = 0, \ \ \forall\ t < t_s(h) \text{ or } t > t_d(h),\ \forall\ h \tag{3.11j}$$

$$x_{ht}^3 \geq x_{h(t-1)}^3, \ \forall t \in [t_s(h), t_s(h) + D(h) - 1],\ \forall\ h \tag{3.11k}$$

$$x_{ht}^3 \geq x_{h(t-1)}^3 - x_{h(t-D(h))}^3, \ \forall t \in [t_s(h) + D(h), t_d(h) - D(h) + 1],\ \forall\ h \tag{3.11l}$$

$$x_{ht}^3 \leq x_{h(t-1)}^3, \ \forall t \in [t_d(h) - D(h) + 2, t_d(h)],\ \forall\ h \tag{3.11m}$$

$$U_{v\,ht}^{\ 3} = c_{v\,h}^{\ 3} x_{ht}^3, \ \ \forall\ t,\ \forall\ h \tag{3.11n}$$

$$x_{ht}^a \in \{0,1\}, \ \ \forall\ t,\ \forall\ h,\ \forall\ a. \tag{3.11o}$$

In a second step, comfort utilities are optimized while guarantying achievable vital utility determined by the first step. The corresponding MILP problem is formulated by equations (3.12) where ${U_v}^M$ denotes the solution of the problem (3.11).

$$\max_{X_{ht}^a, x_{ht}^a} \sum_{t=1}^{t_M} \sum_{h=1}^{H} \sum_{a=1}^{A} U_{c\,ht}^{\;a} \tag{3.12a}$$

s.t.

$$\sum_{t=1}^{t_M} \sum_{h=1}^{H} \sum_{a=1}^{A} U_{v\,ht}^{\;a} \geq U_v^{\;M} \tag{3.12b}$$

$$\sum_{h=1}^{H} \sum_{a=1}^{A} X_{ht}^a \leq C(t), \;\; \forall\, t \tag{3.12c}$$

$$P_m^a(h) x_{ht}^a \leq X_{ht}^a \leq P_M^a(h) x_{ht}^a, \;\; \forall\, t, \,\forall\, h, \,\forall\, a \tag{3.12d}$$

$$U_{v\,ht}^{\;1} = \pi_v^1(h,t) x_{ht}^1, \;\; \forall\, t, \,\forall\, h \tag{3.12e}$$

$$U_{c\,ht}^{\;1} = \frac{\pi_c^1(h,t)}{P_M^1(h)} X_{ht}^1, \;\; \forall\, t, \,\forall\, h \tag{3.12f}$$

$$T_{ht} = T_{h(t-1)} + F(h) X_{ht}^2 + G(h)(T_e(t) - T_{h(t-1)}), \;\; \forall\, t, \,\forall\, h \tag{3.12g}$$

$$T_{ht} \leq T_M(h), \;\; \forall\, t, \,\forall\, h \tag{3.12h}$$

$$U_{v\,ht}^{\;2} \leq \frac{\pi_v^2(h,t)}{T_m(h)} T_{ht}, \;\; \forall\, t, \,\forall\, h \tag{3.12i}$$

$$U_{v\,ht}^{\;2} \leq \pi_v^2(h,t), \;\; \forall\, t, \,\forall\, h \tag{3.12j}$$

$$U_{c\,ht}^{\;2} \leq \frac{\pi_c^2(h,t)}{T_P(h) - T_m(h)} (T_{ht} - T_m(h) y_{ht}), \;\; \forall\, t, \,\forall\, h \tag{3.12k}$$

$$U_{c\,ht}^{\;2} \leq \pi_c^2(h,t) y_{ht}, \;\; \forall\, t, \,\forall\, h \tag{3.12l}$$

$$x_{ht}^3 = 0, \;\; \forall\, t < t_s(h) \text{ or } t > t_d(h), \,\forall\, h \tag{3.12m}$$

$$x_{ht}^3 \geq x_{h(t-1)}^3, \; \forall t \in [t_s(h), t_s(h) + D(h) - 1], \,\forall\, h \tag{3.12n}$$

$$x_{ht}^3 \geq x_{h(t-1)}^3 - x_{h(t-D(h))}^3, \; \forall t \in [t_s(h) + D(h), t_d(h) - D(h) + 1], \,\forall\, h \tag{3.12o}$$

$$x_{ht}^3 \leq x_{h(t-1)}^3, \; \forall t \in [t_d(h) - D(h) + 2, t_d(h)], \,\forall\, h \tag{3.12p}$$

$$U_{v\,ht}^{\;3} = c_{v\,h}^{\;3} x_{ht}^3, \;\; \forall\, t, \,\forall\, h \tag{3.12q}$$

$$U_{c\,ht}^{\;3} = \frac{t_M \;\, (t_d(h) - t)}{D(h)((t_d(h) - t_s(h)) - \frac{D(h)-1}{2})} x_{ht}^3, \;\; \forall\, t \in [t_s(h), t_d(h)], \,\forall\, h \tag{3.12r}$$

$$U_{c\,ht}^{\;3} = 0, \;\; \forall\, t \notin [t_s(h), t_d(h)], \,\forall\, h \tag{3.12s}$$

$$x_{ht}^a \in \{0,1\}, \;\; \forall\, t, \,\forall\, h, \,\forall\, a. \tag{3.12t}$$

Chapter 4

Fine-grained centralized control

In the previous chapter, we proposed a generic framework that allows to define utility functions for any appliance. We showed how this framework can be used to design an ideal planner that optimizes appliances scheduling by targeting the maximization of users utility under available capacity constraints. To deal with appliances having different criticality, we advised to consider a lexicographic ordering of utility functions. This allows us to define a first type of fairness between appliances tied to the utility model. However, a second type of fairness can be defined which is explicitly enforced by the control scheme to balance utility perceived by each user. As discussed previously, we will be interested in three families of control schemes: centralized, partially distributed and distributed. In this chapter, we focus on a centralized approach in which we seek to reach an ideal planner's performance. The other two architectures are the subject of the next chapters. For the proposal of DR schemes, we consider the two types of fairness (utility model dependent and scheme dependent). Proposed schemes can produce ahead-of-time or real-time decisions. We consider exact resolution and approximations and discuss decision quality and optimality.

The chapter is organized as follows: Section 1 presents motivation. Section 2 discuss related work. Specific solution framework for the proposed family of DR solutions is presented in Section 3. Proposed control schemes and corresponding formulation model are introduced in Section 4 and Section 5 respectively. Allocation problem difficulty is discussed in Section 6. In this section, we propose several ways to cope with complexity along with approximation algorithms to solve the allocation problem. In Section 7, we present a numerical analysis of the model which outcomes provide insight on the behavior of the proposed control schemes. Conclusions are presented in Section 8.

1 Motivation

Fine-grained direct control provides means for a flexible management of grid resources. It allows to cope with grid operation constraints while minimizing the impact on end users. The legacy grid infrastructure has severe limitations for the fine control of power allocation. This is due to the lack of fine grained information and adequate control equipments. Indeed, control is mainly coarse-grained on the level of distribution network feeders. In this chapter, we consider advanced solutions, like the one presented in [50], which enable controlling the demand at fine granularity (at a per appliance level).

The design of the DR mechanisms implemented through functional groups presented in Section 1.1 targets energy efficiency as well as satisfying users' expectations. Reaching these objectives is challenging because, on the one hand, those two global objectives could be contradictory and, on the other hand, theoretic optimal solutions may not be achieved (e.g. for complexity reasons). We will be interested in finding a satisfying trade-offs between performance and complexity.

We can summarize our problem statement as follows: under the hypotheses that an advanced architecture enabling demand control at fine granularity is deployed and that relevant information is monitored at user premises, we search demand control mechanisms for energy efficiency and users satisfaction taking into account different constraints of scalability and fairness.

2 Related work

Several approaches for fine grained control can be found in the literature. Some directly take decisions based on power profile of appliances (e.g., [6, 26, 54]) while others model user's quality of experience (e.g., [86, 55, 97, 50]).

In some proposals like [6, 26], authors suppose the capability of scheduling all appliances over the DR period and focus on modeling shifting capabilities of some loads. In [54], a different approach is supposed where decisions are produced such that a maximum interruption duration constraint is satisfied. Such approaches are clearly not suitable for our problem since we do not make any assumptions on the possible value of available capacity. Indeed, controlled appliances may not be all scheduled and a maximum interruption duration cannot be guarantied when available capacity is extremely scarce.

Utility functions allow to guide control decisions by intrinsically expressing the flexibility and relevance of controlled appliances for users. They were previously used in the literature to measure social welfare when designing pricing schemes and are generally considered as (convex) functions of power (e.g., [86]). In [55], authors propose DR schemes that are capable of taking into account the utility of appliances modeled based on their usage features. As stated in Section 2, our utility modeling approach is similar and extends the one presented in [55] for utility functions definition.

For direct load control, some proposals use utility functions to design centralized control schemes. However, in most cases, approximation algorithms optimality are not discussed. For instance, authors in [97] propose a customer reward scheme that incentivizes users to accept direct control of loads. They propose a greedy algorithm that maximizes utility of appliances based on the values they declare for each time slot. In [50], authors consider advanced solutions which enable controlling the demand at fine granularity while taking into account appliance's criticality. In the present chapter, we extend their proposal by introducing different ways of dealing with fairness and discussing different decision timescales.

3 System architecture

We suppose a framework where the aggregator (DE_a) has to control a set of appliances in order to limit consumption to a certain desired capacity over a known period of time. We call this period of time a DR period.

In this chapter, we will be focusing on centralized solutions where decisions are only taken by the DE_a. In this case, the role of DE_h is to collect relevant home information and send it to the DE_a. They also relay control decision issued by the DE_a to the appliances. Such fully centralized architecture is illustrated by Figure 4.1.

To reach its goal DE_a should implement schemes that: i) can control and interact with users' appliances, ii) can take into consideration data monitored at users' premises, like presence and temperature and iii) based on this information, optimizes the total provided utility (e.g. at a neighborhood scale) under fairness constraints. Fairness definition relies on the control scheme and is discussed in Section 4.2.

4 Proposed control schemes

In our DR proposals, we make several design choices namely related to the control target, fairness considerations and decision time scale. Indeed, we focus on proposing services that can provide some quality of service guaranties for users. So, even though we do not discuss tariff schemes that can be associated to them, we suppose that such tariffs should reflect the service level guarantied. To introduce fairness between users, we previously discussed a definition that relies on a utility model where different criticality levels are introduced (see Section 3). However, we may envision a system considering only one criticality level in which case fairness needs to be introduced as constraints of the control scheme. In this section, we propose control schemes that provide fairness guaranties between homes. They are supposed to analyze trade-offs between the two possible definitions of fairness. Finally, we analyze two times-scale for producing control decisions namely ahead-of-time and real-time. While the first is considered to allow resource provisioning and optimal control, the second one allows greater reactivity.

4.1 Maximum Utility

The Maximum Utility schemes suppose DE_a has full information and decide how much power to allocate to each appliance in order to maximize the total utility. Total utility is defined by the sum of the utilities of each appliance at each home (see Chapter 3).

4.2 Fairness constraints

We consider schemes similar to Maximum Utility in the sense that DE_a has full information. However, their objective is to maximize the total utility in a restricted domain where the minimum of the utilities of all homes is maximized (see Section 5 for a formal definition). So, for this scheme, we try to maximize the utility of the most disadvantaged home(s) (max-min).

We consider such schemes in order to introduce additional fairness guaranties between homes to supplement fairness that can be introduced by the utility function model (through the lexicographic ordering). In fact the specific fairness definition we suppose here can seem in particular appropriate for the case of homogeneous homes (i.e., homes with similar characteristics). For the case of heterogeneous homes, depending on different considerations that we will not treat here, it can be argued that a scheme where max-min is applied to the utility normalized depending on homes characteristics (e.g., normalizing while taking into account the number of inhabitants or square meters) would be fairer. In addition, lexicographic ordering of utility functions may provide more flexibility by allowing to introduce fairness between appliances.

Notice that in this case we maximize the minimum (max-min). However, this is not max-min fairness; nevertheless, the minimum utility provided to a given user will be the same as in max-min fairness. We do not introduce max-min fairness because max-min fair solutions of the problem we define may not exist, as can be understood from the formal problem stated in Section 5 when utility function model is not restricted to convex functions.

4.3 Decision time

Control schemes can be studied considering two types of approaches, namely *ahead-of-time* and *real-time* (see Chapter 2, page 12). For this chapter, let us precise our definitions of ahead-of-time and

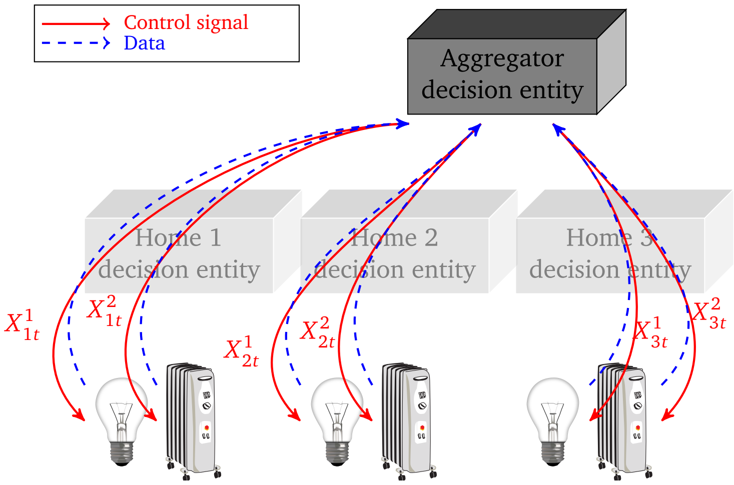


Figure 4.1 – Centralized control

real-time. Supposing time is slotted, we call *ahead-of-time* those approaches where the decisions are taken at the beginning of a time period for its whole duration. We call *real-time* approaches those where for a given time period of t_M time slots, control decisions are taken at each time slot based on knowledge of exogenous variables and system parameters (like available capacity) of the specific time slot. In this case, the way exogenous variables (e.g., temperature) will evolve in future time slots is not taken into account.

The distinction ahead-of-time/real-time has a dual interest.

First, it allows to study two situations of generation scarcity corresponding to predictable and unpredictable events. So, the DE_a at the DSO or aggregator may be able to forecast the starting time and duration of a starvation time period. In this case, the power allocation may be done at the beginning of the time period. Hence, scheduling can be decided on the whole duration. Such situation typically occurs in case of a crisis due to extreme weather conditions. Otherwise, the starvation periods are unknown and decisions can only be taken based on instantaneous available information.

Secondly, optimizing the system for the whole period under studied constraints that take into account the dynamics of different variables (like the temperature in the house) may cause scalability issues due to complexity. A slot per slot optimization may enhance scalability at the price of a suboptimal allocation. Thus, analyzing both approaches allows to measure the gap between the performance of decisions taken ahead-of-time and those taken in real-time.

5 Model

We propose to model all the above-presented schemes both in cases of ahead-of-time and real-time optimization. The target of the system is to offer the highest possible quality of experience while complying with system capacity constraints $C(t)$ for all time instants t.

Next, in line with the framework proposed in the previous chapter, we present a set of optimization problems that expresses the previous sentence and the system constraints in a formal language for each of the control schemes we have introduced. We denote by t_M the duration of the DR period in number of time slots, by H the number of homes and by A the number of appliances per home. We remind that the main notation of this thesis is summarized in Table 3.1, page 35.

5.1 Maximum Utility, ahead-of-time

We assume that the DE_a has a complete view on the system and produces its decisions ahead-of-time based on utility maximization objective. We call this scheme Global Maximum utility and we name it GM. The centralized global optimization for this scheme is essentially the one exposed in Section 3.2, page 37 formulated by equations (4.1).

$$\max_{X^a_{ht}, x^a_{ht}} \sum_{t=1}^{t_M} \sum_{h=1}^{H} \sum_{a=1}^{A} U^a_{ht} \tag{4.1a}$$

s.t.

$$\sum_{h=1}^{H} \sum_{a=1}^{A} X^a_{ht} \leq C(t), \quad \forall\, t \tag{4.1b}$$

$$P^a_m(h) x^a_{ht} \leq X^a_{ht} \leq P^a_M(h) x^a_{ht}, \quad \forall\, t, \forall\, h, \forall\, a \tag{4.1c}$$

$$x^a_{ht} \in \{0,1\}, \quad \forall\, t, \forall\, h, \forall\, a. \tag{4.1d}$$

The problem (4.1) can be solved by the DE_a if all the informations about appliances and their utility functions are transmitted by each home DE_h. DE_a can then compute an optimal solution and notify homes accordingly.

Decision variables in this case are variables x^a_{ht} and X^a_{ht}. Binary variables x^a_{ht} correspond to turning ON (i.e., $x^a_{ht} = 1$) or OFF (i.e., $x^a_{ht} = 0$) appliance a at home h on time slot t. If appliance is turned ON, power allocation X^a_{ht} can take values between a minimum value $P^a_m(h)$ and a maximum value $P^a_M(h)$ (see equation (4.1c)).

In this formulation as well as the next ones, we do not explicit the constraints for utility functions variables U^a_{ht} as they depend on the specific appliances considered. In addition to constraints related to utility variables, some appliances may also require additional constraints related to power (x^a_{ht} and X^a_{ht}). These constraints can be added, for instance, to force a uninterrupted operation. Examples of a complete formulation of constraints can be found in Section 4.

5.2 Maximum Utility, real-time

The maximum utility in a real-time approach is denoted by RT. Although the decisions are taken here per time slot, the dynamics of the system are still considered. For example, temperature is calculated based on the modeled temperature dynamics, exogenous variables and previously taken decisions. This enables to compare the performance of this scheme with the previous one. Indeed, although designed for usage in real-time, this scheme can also be used ahead-of-time, as it reduces the computation time and therefore increases scalability compared with the previous one (of course, such approach will lead to poorer performance). The constraints are the same as in the previous case and we have now t_M problems to solve consecutively given by equations (4.2).

$$\max_{X^a_{ht}, x^a_{ht}} \sum_{h=1}^{H} \sum_{a=1}^{A} U^a_{ht} \text{ for } t = 1..t_M. \tag{4.2}$$

5.3 Maximum Utility under fairness constraints, ahead-of-time

Now, we consider ahead-of-time approach for maximum Utility under fairness constraints that we call Fair Global Maximum utility (FGM). This scheme is made of two stages.

In the first stage, we compute the domain of the variables where the minimum utility that users may perceive, is maximized. This problem is formulated by equation (4.3) where U expresses minimum utility that we want to maximize.

$$\max_{X_{ht}^a, x_{ht}^a} \; U \tag{4.3a}$$

s.t.

$$\sum_{t=1}^{t_M} \sum_{a=1}^{A} U_{ht}^a \geq U, \;\; \forall \, h \tag{4.3b}$$

$$\sum_{h=1}^{H} \sum_{a=1}^{A} X_{ht}^a \leq C(t), \;\; \forall \, t \tag{4.3c}$$

$$P_m^a(h) x_{ht}^a \leq X_{ht}^a \leq P_M^a(h) x_{ht}^a, \;\; \forall \, t, \, \forall \, h, \, \forall \, a \tag{4.3d}$$

$$x_{ht}^a \in \{0, 1\}, \;\; \forall \, t, \, \forall \, h, \, \forall \, a. \tag{4.3e}$$

Let us denote by U^M the solution of the previous problem. This tells us that there exists a solution where all users have a utility at least equal to U^M.

In the second stage, we search for the best solution that guaranties U^M for all users. This problem is expressed by equations (4.4) which maximizes the global utility inside the domain computed by the previous stage.

$$\max_{X_{ht}^a, x_{ht}^a} \sum_{t=1}^{t_M} \sum_{h=1}^{H} \sum_{a=1}^{A} U_{ht}^a \quad \text{s.t.} \tag{4.4a}$$

s.t.

$$\sum_{h=1}^{H} \sum_{a=1}^{A} X_{ht}^a \leq C(t), \;\; \forall \, t \tag{4.4b}$$

$$p^a(h) x_{ht}^a \leq X_{ht}^a \leq P^a(h) x_{ht}^a, \;\; \forall \, t, \, \forall \, h, \, \forall \, a \tag{4.4c}$$

$$\sum_{t=1}^{t_M} \sum_{a=1}^{A} U_{ht}^a \geq U^M, \;\; \forall \, h \tag{4.4d}$$

$$x_{ht}^a \in \{0, 1\}, \;\; \forall \, t, \, \forall \, h, \, \forall \, a. \tag{4.4e}$$

Remark: this scheme might give a different solution compared to GM, even if similar homes are considered. Let us suppose an example where three homes participate to the scheme (h=1,2 or 3). These homes each have 2 controlled appliances each belonging to a class (either a=1 or a=2). Suppose an optimization period of 1 time slot where available capacity $C(t) = 4$. Suppose $p^a(h) = P^a(h) \; \forall h \; \forall a$. Let us take power values $p^1(h) = 1 \; \forall h$ and $p^2(h) = 2 \; \forall h$, and utility values $u^1(h) = 1 \; \forall h$ and $u^2(h) = 3 \; \forall h$ perceived when operating appliances of the first class and the second class respectively. We can see that, using GM, the total maximum achievable utility is 6 corresponding to scheduling two appliances of the second class in two homes. As for FGM, total maximum utility is equal to 5 corresponding to scheduling two appliances of class 1 in two homes and an appliance of class 2 in the third home.

5.4 Maximum Utility under fairness constraints, real-time

In a similar way to RT, we can derive the real-time approach of maximum utility under fairness constraints related to improving the worst case scenario for homes. We call it FRT.

We have t_M successive problems each made of two steps to solve consecutively. In the first step, the value U_t^M is found for a given t as a solution of equation (4.5).

$$\max_{X_{ht}^a, x_{ht}^a} \; U_t \tag{4.5a}$$

s.t.

$$\sum_{a=1}^{A} U_{ht}^a \geq U_t, \;\; \forall \, h \tag{4.5b}$$

Based on U_t^M, appliances allocations are determined by solving the problem formulated by equation (4.6)

$$\max_{X_{ht}^a, x_{ht}^a} \sum_{h=1}^{H} \sum_{a=1}^{A} U_{ht}^a \quad \text{s.t.} \tag{4.6a}$$

s.t.

$$\sum_{a=1}^{A} U_{ht}^a \geq U_t^M, \ \forall\, h. \tag{4.6b}$$

6 Solving the problem

To solve the optimization problems, we can either attempt an exact resolution or search for an efficient heuristic. In this section, we discuss ways to implement both resolution methods. Indeed, first, we propose techniques that can enable exact resolution by alleviating the size of the allocation problem. These techniques consist in choosing adequate resolution parameters. Second, for a given parameter choice, we propose greedy heuristics and discuss their optimality.

6.1 Exact resolution

In this section, we discuss the complexity of solving the allocation problem and advise techniques to enhance the scalability. Indeed, one can use MILP solvers to compute control decisions for the previously formulated problems such as CPLEX [45]. To improve resolution time, we propose ways to limit the number of variables and constraints of the problem. We focus on the case where system parameters can be predicted in advance and a solution for the Global Maximum utility (GM) scheme is desired.

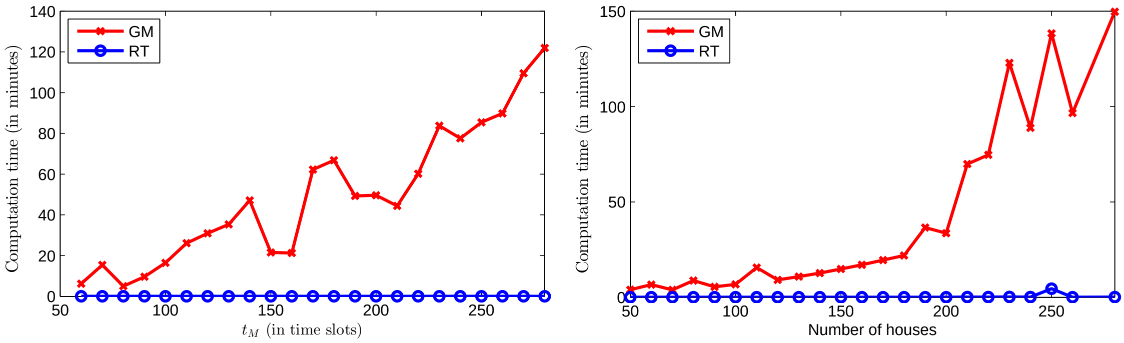
6.1.1 $\mathcal{NP}$-completeness of the allocation problem

In this section, we show that the global allocation problem (which is closely related to generalized assignment problems) is $\mathcal{NP}$-complete. To do this we shall show that the 0-1 single KNAPSACK Problem ([33, p. 247]) can be transformed to (a simplified version of) GM. 0-1 KNAPSACK Problem is known to be $\mathcal{NP}$-complete and is as follows:

PROBLEM: KNAPSACK
INSTANCE: A finite set of items A. Each item $a \in A$ has weight $w(a) \in \mathbb{Z}^+$ and a profit $u(a) \in \mathbb{Z}^+$. The knapsack has positive capacity C and a value goal K.
QUESTION: Is there a subset $A' \subseteq A$ such that:
$\sum_{a \in A'} w(a) \leq C$ and $\sum_{a \in A'} u(a) \geq K$.

Proposition 1. *Problem GM is $\mathcal{NP}$-complete.*

Proof. Suppose an instance of KNAPSACK that takes as input the set of items A, capacity C and a utility goal K. It can be transformed to a GM formulated by equation (4.1) by taking one time slot ($t_M = 1$), a capacity $C(t) = C$ and, for each item $a \in A$, minimum and maximum bounds of equation (4.1c) equal to $w(a)$ and a utility equal to $u(a)$ if the item is scheduled. If we can solve GM, we can solve KNAPSACK after the straightforward transformation described previously. The transformation is linear (the inputs of KNAPSACK are directly injected into GM), which proves the NP-completeness of GM. □



(a) Time as a function of the optimization duration

(b) Time as a function of the number of houses

Figure 4.2 – Computation time of maximum utility problems

6.1.2 Improve scalability

The number of variables in the optimization problem increases with the number of houses and appliances, and the number of time slots. Choosing time slot duration, optimization period, appliances or houses that will take part in the optimization are possible ways to control the size of the problem that need to be solved. Indeed, a reduced size of the optimization problem will allow a faster resolution time. This is illustrated in Figure 4.2 which shows the computation time required to solve GM and RT problems supposing different optimization duration and different number of homes. For presented results, we suppose a homogeneous scenario where all homes belong to the same class (Class 1) and a utility function model where vital and comfort components are supposed (see Section 7 for details on numerical settings). In Figure 4.2a, we fix all parameters and change optimization period. We can observe that global maximum computation time increases exponentially with the optimization duration while RT scheme have a significantly lower computation time. Similarly, for Figure 4.2b, we fix all the parameters and analyze computation time for different values of the available capacity and the number of homes. Suppose H the number of homes. We take a constant capacity C over the optimization period which satisfies $C = 1000\ H$. We can observe that GM computation time is significantly higher than RT and has an exponential increase in the number of homes.

However, such simplifications impact the quality of the resulting solution. So, a trade-off is required between resolution time and solution quality.

We now discuss how the problem size can be reduced by proposing ways to reduce the number of time slots and the number of appliances involved in the optimization.

Dealing with time

Reducing the number of time slots can be achieved in three ways: increasing the time slot duration τ, aggregating decisions on time slots (i.e., take time slices that are a multiple of τ) or reducing optimization period t_M.

Time slot duration

Increasing time slot duration τ is a straightforward way to reduce the number of decision variables since for a given interval of time, the number of considered slots will decrease. However, this technique affects the time flexibility of all controlled appliances. Hence, it induces degradation in the quality of the solution.

Aggregating time slot for appliances
Another way to reduce the number of variables by reducing the number of time slots is to aggregate slots. Supposing a fixed slot duration τ, we can define for each appliance a decision step that is a multiple of τ. Then, the number of variables related to an appliances (e.g., power allocation) will depend on this chosen step. So instead of having decisions over t_M time slots, the number of variables related to a given appliance a will depend on t_M/t_a for a decision step equal to $t_a\tau$. This is different from the previous case since decision steps can be different for different appliances.

This technique can also be used to reduce the number of constraints of the optimization problem. Indeed, it allows for example to enforce a minimum duration of operation without having to add constraints on power allocation.

Although having appliance specific slot duration can help reduce the size of the problem, it has several shortcomings. First, for a given appliance a, minimum ON and OFF periods will be equal to the decision step ($t_a\tau$). Second, since power is constant on a decision step, taking a longer step will reduce the flexibility that can be provided by the appliance (i.e., opportunities to power ON and OFF or modulate power consumed by the appliance).

Optimization period
One of the main advantages of solving the optimization problem over the whole optimization duration is to be able to have a long term overview of the utility of appliances. As shown in Figure 4.2, solving the problem for time slots consecutively (i.e., real-time schemes) can help to reduce the resolution time of the problem. This is the extreme case of optimization period reduction. A generalization is also possible. Indeed, one can suppose a resolution over a sliding time window. In this case, the problem is solved supposing on a shorter time period (corresponding to the window size). Then decision variables values are fixed for the time slots that will not be part of the next optimization after shifting the optimization window. The optimality of such resolution will depend on the size of the optimization window, the number of slots we are shifting the optimization window and the types of appliances that are taking part of the optimization problem. An example that illustrates a case where this might be suboptimal is when an appliance that cannot be interrupted is supposed. Indeed, if it is operated at a time slot based on the optimization outcome over a time window. Then , for the next optimization time window, if the appliance is still operating, it may prevent an appliances of higher criticality from being operated. This is the case when available capacity is not enough for fulfilling both demands.

Optimization by type
Consecutive optimization by type of usage (i.e., interactive, background or program-based) is a first possible way to reduce the number of involved appliances in the optimization. Supposing an ordering of the three types of usage, this consists in scheduling appliances belonging to the first chosen type and updating available remaining capacity. then, the same procedure is repeated for the next two types considered based on the decided order. Using this technique allows to derive control policies that depend on the type of load. However, this does not provide any guaranties on the global optimality of the allocation.

A resolution by type can be implemented in a homogeneous setting were appliances have the same criticality and priority. However, its implementation becomes difficult when a heterogeneous case is supposed. Indeed, difficulty arises for instance when users have different criticality and priority for each type. So, it becomes challenging to decide in which order the different types will be considered for optimization.

Optimization by house
Reducing the problem size can be done by decomposing the central problem into smaller problems that involve one home or a smaller group of homes. In this case, a master problem will coordinate subproblems solved using information on the level of homes. A technique based on this approach is proposed and discussed in the next chapter.

6.2 Greedy heuristics

In this section, we focus on proposing efficient greedy heuristics to solve the Global Maximum utility problem. We also discuss their optimality.

6.2.1 Algorithms

Depending on the control time scale, two approaches can be considered. The first can take decisions on scheduling appliances at any time slot without any ordering restriction. We call this approximation a "global greedy" algorithm. This corresponds to a model where ahead-of-time decisions are taken. The second approach requires scheduling to be done for each slot in chronological order. However, this is different from real-time schemes proposed previously. Indeed, in this case, the dynamics of system variables in future time slots are modeled and system parameter (e.g., available capacity) can be known. We call this algorithm by "Real-time greedy" to refer to decision time scale.

For all heuristics, power needs to be discretized to reduce allocation options. Indeed, since appliances may have adjustable power input between a minimum power required to operate and a maximum power, we introduce the increment ΔP. This increment defines a minimum power increase that can be added to an appliance's allocation.

The global greedy heuristic computes the cost (in terms of power required) and the value of operating or increasing power allocation (by ΔP) for each appliance at each time slot of the optimization period. For each appliance a at home h, we define the additional value denoted by v_{ht}^{a} of providing power of w_{ht}^{a} at time t. Computing the value v_{ht}^{a} is discussed in Section 6.2.2. Power w_{ht}^{a} represents either the power needed to start the appliance at time slot t (i.e., $P_{m}^{a}(h)$) or additional power that can be added to the appliance's allocation (i.e., ΔP). Indeed, some appliances (e.g., washing machines) require a minimum operation duration. This means that starting the appliance at time t will produce power consumption at subsequent time slots.

A list L_1 is defined with "items" from appliances that can be scheduled or for which power allocation can be increased depending on system available capacity ($C(t)\ \forall t$) or other homes constraints (e.g., appliances maximum power, maximum temperature). An item has a weight w_{ht}^{a} and a value v_{ht}^{a}. It corresponds to an appliance a in home h that can be operated at time t. The list L_1 also specifies the efficiency v_{ht}^{a}/w_{ht}^{a} of allocating power to each appliance a in each home h at a time slot t. At each iteration of the global greedy heuristic, the most efficient appliance is scheduled at the most profitable time slot (i.e., the highest $v_{ht}^{a}/w_{ht}^{a}\ \forall a,\ \forall\ h,\ \forall\ t$). The list of items is updated and the selection process is repeated until no further allocations can be done. We note that appliances already scheduled and have reached its maximum power allocation will be taken out of the list of appliances that need to be allocated (i.e., $v_{ht}^{a} = 0$). In addition, some appliances can only be scheduled in certain time slots (e.g., a washing machine can be scheduled starting from the earliest start time and its operation should end by a deadline). The global greedy heuristic can be described by the Algorithm 2.

The real-time greedy heuristic is similar to the global greedy one. However, decisions are made for each time slot consecutively. Indeed, a list L_2 is defined and populated with items that can be scheduled at a given time slot. For this time slot, the most cost effective item is chosen and corresponding

Algorithm 2 Global greedy algorithm

procedure GGREEDY
  Step 0:
  Initialize decision variables $X^a_{ht} = 0, x^a_{ht} = 0, \ \forall\ t, \forall\ h, \forall\ a;$
  Initialize items each with a weight $w^a_{ht} = P^a_m(h)$ and a value $v^a_{ht}, \ \forall\ t, \forall\ h, \forall\ a;$
  Initialize the list L_1 with items from appliances that can be scheduled based on capacity $C(t)\ \forall\ t$.
  while $L_1 \neq \emptyset$ **do**
    Step 1: Choose the most cost effective item in L_1 (say corresponding to appliance a_i in home h_j at time slot t_k with the highest $v^{a_i}_{h_j t_k}/w^{a_i}_{h_j t_k}$).
    Step 2: Assign power $w^{a_i}_{h_j t_k}$ to appliance a_i in home h_j by updating decision variables $X^{a_i}_{h_j t}$ and $x^{a_i}_{h_j t}$ for $t \geq t_k$.
    Step 3: Update the list L_1 by updating the weight and value of items that can be scheduled based on remaining capacity $(C(t) - \sum_{h=1}^{H}\sum_{a=1}^{A} X^a_{ht}\ \forall t)$ and removing items that cannot be scheduled.
  Output power allocations X^a_{ht} and $x^a_{ht}, \ \forall\ t, \forall\ h, \forall\ a$.

appliance is operated based on its power allocation (i.e., item's weight). After each allocation decision, the list L_2 is updated and the process is repeated until no further allocations can be done based on system available capacity and appliances' operation constraints at the supposed time slot. Once the allocation process over a time slot is completed, we move to the next time slot for which the same selection process is applied. This is done for all time slots till the end of the optimization period.

The real-time greedy heuristic can be described by the Algorithm 3.

Algorithm 3 Real-time greedy algorithm

procedure RTGREEDY
  Step 0:
  Initialize decision variables $X^a_{ht} = 0, x^a_{ht} = 0, \ \forall\ t, \forall\ h, \forall\ a;$
  for $t = 1$ to t_M **do**
    Initialize items each with a weight $w^a_{ht} = P^a_m(h)$ and a value $v^a_{ht}, \forall\ h, \forall\ a;$
    Initialize the list L_2 with items from appliances that can be scheduled based on system capacity $C(t)$.
    while $L_2 \neq \emptyset$ **do**
      Step 1: Choose the most cost effective item in L_2 (say corresponding to appliance a_i in home h_j with the highest $v^{a_i}_{h_j t}/w^{a_i}_{h_j t}$).
      Step 2: Assign power $w^{a_i}_{h_j t}$ to appliance a_i in home h_j by updating decision variables $X^{a_i}_{h_j t_k}$ and $x^{a_i}_{h_j t_k}$ for $t_k \geq t$.
      Step 3: Update the list L_2 by updating the weight and value of items that can be scheduled based on remaining capacity $(C(t) - \sum_{h=1}^{H}\sum_{a=1}^{A} X^a_{ht})$.
  Output power allocations X^a_{ht} and $x^a_{ht}, \ \forall\ t, \forall\ h, \forall\ a$.

6.2.2 Specifics per type

We now discuss how this algorithm can be implemented for each type of usage to derive the additional value v^a_{ht} from allocating certain power w^a_{ht} to an appliance a. Regardless of the type, each appliance a

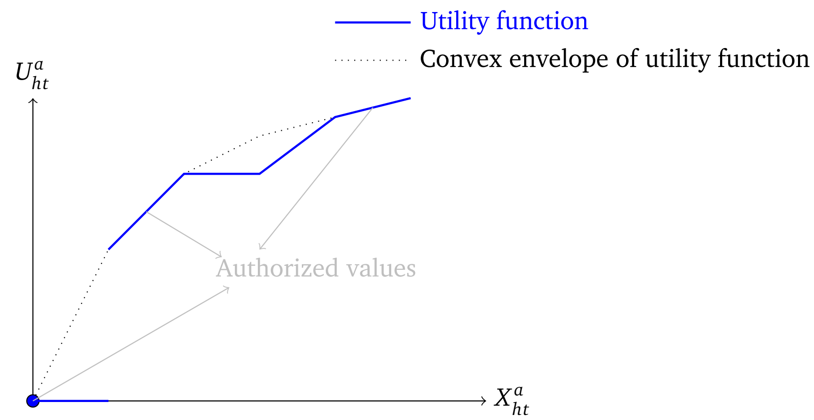


Figure 4.3 – Utility function and convex envelope

will be represented by $1 + \lceil (P^a_M(h) - P^a_m(h))/\Delta P \rceil$ items that can be packed at each time slot. The first item that need to be packed is the one corresponding to minimum power required for the appliance to operate.

Interactive demands

For interactive loads since utility functions are usually independent between time slots, the general problem can be seen as a single knapsack problem for each time slot if no additional constraints on power are supposed. This is the case of the lighting example we take (see Section 4.1). In this case, both global greedy and real-time greedy will render the same solution.

Indeed, the value of each item depends on power already allocated and can be directly derived from the utility function at the time slot in which an item can be scheduled. So, v^a_{ht} for an item of an interactive appliance a is equal to the difference between the value corresponding to packing the item (i.e., $X^a_{ht} + w^a_{ht}$) and the value corresponding to already scheduled items of a (i.e., X^a_{ht}). These values are taken from the convex envelope of the utility function U^a_{ht}. As presented in Figure 4.3, a utility function U^a_{ht} may be non-convex in power X^a_{ht}. This function is represented by the full line while the dotted line represents its convex envelope. In this case, taking values from the utility function may prevent from capturing the utility opportunity of scheduling a chunk of power w^a_{ht} for a certain appliance a on future packing decisions of the same appliance. This can be illustrated by Figure 4.3 when scheduled power X^a_{ht} and power $X^a_{ht} + w^a_{ht}$ have utilities corresponding to the third linear piece of the utility function. Indeed, taking the values of the utility function will show little additional value for scheduling even if utility can be significantly increased when power allocation is further increased. Some pieces of the utility function can be authorized to consider since they coincide with the convex envelope. If all items that can be generated by an appliance are supposed, taking values from the convex envelope allows to schedule these items in the correct order. Suppose appliance a may have up to k items to schedule (i.e., the first item has a weight equal to minimum power $P^a_m(h)$ and the rest has weights equal to ΔP). A convex value model will render items' efficiency decrease with increased power allocation. This is due to the fact that the slope of a convex utility function decreases with power. This makes the first item more attractive to schedule than subsequent ones and so on.

Background loads

For background loads, when the impact of variables spans multiple time slots, time dependence between decisions needs to be taken into account. In the following, we elaborate using the example of

heating (see Section 4.2). When considering heaters, heating at time slot t will generate additional temperature that will affect subsequent time slots (due to inertia). So the value of heating at a time slot t is equal to the utility value generated by heating at t and additional utility values gained from temperature increase in subsequent time slots.
The value of operating a heater a at particular time slot t can be written: $v^a_{ht} = \sum_{t'=t}^{t_M} \Delta U^a_{ht'}$ where $\Delta U^a_{ht'}$ denotes the additional gain in utility due to the temperature increase at time slot t' generated by heating at time slot t.

We remind that the temperature evolution model is considered to follow the equation below: $T_{ht} = T_{h(t-1)} + F(h)(X^a_{ht} + w^a_{ht}) + G(h)(T_e(t) - T_{h(t-1)})$.

So, heating at time slot t will give temperature T_{ht}. In following time slots, temperature evolves according to equations:
$T_{h(t+1)} = T_{ht} + G(h)(T_e(t+1) - T_{ht})$
$T_{h(t+2)} = T_{h(t+1)} + G(h)(T_e(t+2) - T_{h(t+1)})$, and so on.

Suppose ΔT_{ht} denotes the temperature change at time t due to heating at a time slot $t' \leq t$. If heating is activated at a time slot t, a temperature increase ΔT_{ht} is generated. Based on temperature evolution equation, we can write how temperature will increase due to heating in subsequent time slots by only taking terms that depend on previous temperature:
$\Delta T_{h(t+1)} = \Delta T_{ht}(1 - G(h))$;
$\Delta T_{h(t+2)} = \Delta T_{h(t+1)}(1 - G(h))$ and so on.
Based on these equations, $\Delta T_{ht'}$ (for $t' = t+1$ to t_M) can be expressed based on ΔT_{ht} by equation: $\Delta T_{ht'} = \Delta T_{ht}(1 - G(h))^{(t'-t)}$. We can compute accumulated temperature:
$\sum_{t'=t+1}^{t_M} \Delta T_{ht'} = \Delta T_{ht}(1 - G(h))\frac{1-(1-G(h))^{(t_M-t)}}{G(h)}$.

To compute the value of heating, we consider two approaches. They are differentiated by the way multi-level utility model is dealt with.

In the first approach, we will compute the exact value of heating by comparing utility function values supposing temperature evolution with and without heating. We will denote greedy algorithms following this technique by $GGreedy1$ and $RTGreedy1$ that correspond to a global and a real-time greedy implementation respectively. To compute the value v^a_{ht} of w^a_{ht}, we take the value ΔU^a_{ht} corresponding to the value of a temperature increase of $F(h)w^a_{ht}$. For subsequent time steps (i.e., $t' = t+1$ to t_M), utility increase $\Delta U^a_{ht'}$ is computed as the utility of the temperature increase between the value of temperature evolving from T_{ht} with w^a_{ht} being allocated and the value of temperature evolving starting at t without w^a_{ht} being added to allocation variable X^a_{ht}.

In the second approach, we compute an approximation of the value of heating based on the utility level that needs to be fulfilled at a certain time slot. In the case of vital and comfort utility levels, this will depend on the temperature being above or below minimum tolerated temperature $T_m(h)$. Indeed, with heating at time slot t, if the temperature is above $T_m(h)$, only comfort utility is supposed for additional temperature in subsequent time slots. If the temperature is below $T_m(h)$, vital utility is assumed. This is different from the first approach for which exact utility value is computed. Indeed, even if temperature is above $T_m(h)$ at the time slot where heating can be operated, it can go below $T_m(h)$ in future time slots. For the first approach, this will generate vital utility during time slots in which temperature is below $T_m(h)$ and comfort utility in the rest of the time slots.

For the second approach, to calculate the value of heating v^a_{ht} of w^a_{ht}, we take the value ΔU^a_{ht} corresponding to $\Delta T_{ht} = F(h)w^a_{ht}$ increase in temperature. Based on the added value (vital or comfort) at time t, we compute the corresponding $\sum_{t'=t+1}^{t_M} \Delta U^a_{ht'}$ by replacing $\sum_{t'=t+1}^{t_M} \Delta T_{ht'}$ by the previously obtained expression in ΔT_{ht}. Global and real-time greedy algorithms following this approach are denoted by $GGreedy2$ and $RTGreedy2$ respectively.

Program-based loads

Since Program-based loads have utility functions that are usually coupled with restrictions on operation duration over the DR period, computing a value for an item of this type of appliances is challenging because of the strong dependence between time slots.

Suppose that an item at a time slot t corresponds to starting the appliance operation at time t. One possible way of deciding the value of this item is to take the efficiency over all operation time slots: $\sum_{t'=t}^{t+D(h)-1} U^a_{ht}/(D(h)P^a_M(h))$ where $D(h)$ is operation duration. However, no guaranty can be given on the optimality of such a strategy.

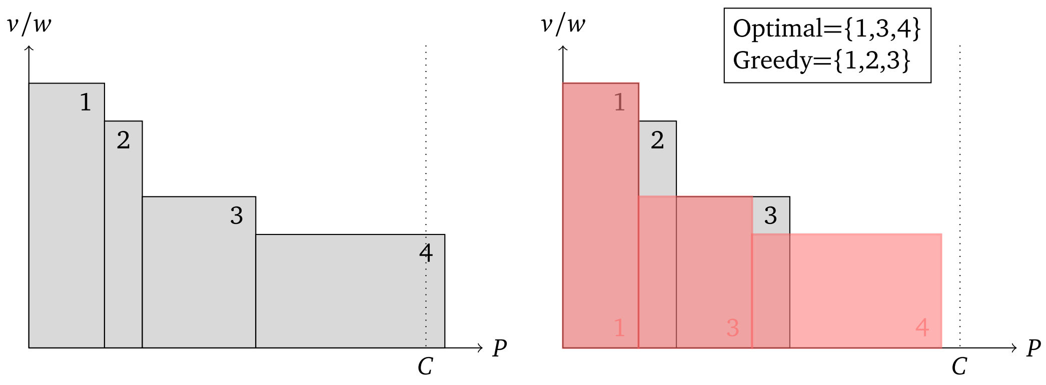
6.2.3 Performance bounds

Suppose X^* is an optimal solution to the optimization problem involving a certain type of appliances. Let us denote by UG the objective value obtained using greedy algorithms based on value increase of scheduling an appliance. In this section , we use *relative value* to denote the utility value increase if an appliance item is scheduled over the item size. We will be interested in estimating the gap between the optimal and allocation based on greedy scheduling presented previously supposing each of the three types of usage we define. Two types of performance guarantees are discussed namely *relative* (multiplicative) and *absolute* (additive) performance bounds. For relative bound, we will be interested in finding ϵ_x such that $UG \geq (1-\epsilon_x)X^*$. Additive performance bound consists in finding ϵ_+ such that $UG \geq X^* - \epsilon_+$.

Interactive demands

As stated in Section 6.2.2, scheduling interactive loads over a period t_M is similar to solving t_M independent knapsack problems. Greedy techniques that take the relative value based on the size of the item are known not to a have a relative bound. Indeed, suppose two items 1 and 2 have a utility u_1 and u_2 and weights w_1 and w_2 respectively. Suppose one and only one of them can be packed because of capacity limitation ($w_1 \leq C$,$w_2 \leq C$ but $w_1 + w_2 > C$). An arbitrarily bad performance can be seen if we suppose $u_2 > u_1$ but $u_2/w_2 < u_1/w_1$, in which case item 1 gets selected by the greedy algorithm. As a numerical example satisfying previous inequalities, suppose $C = 1$, $u_1 = 2\epsilon$, $w_1 = \epsilon$, $u_2 = 1$, $w_2 = 1$. In this case, the greedy algorithm will select item 1 and hence gives a total utility of 2ϵ instead of 1. The selection choice remains unchanged for any value $\epsilon << 1$.

As for the absolute bound, it can be represented by the additive error we can make in each slot. For one slot, we can obtain the error bound by sorting items generated by appliances by decreasing order of relative value depending on the size and packing them following this order. This can be illustrated in Figure 4.4a. In this figure, four items (denoted by 1, 2, 3 and 4) are ordered in decreasing efficiency (i.e., relative value). The x-axis represents the weights of items and the y-axis represents the efficiency. The parameter C represents available capacity at the time slot. In case of convex functions for utility (or convex envelope of a non-convex function), the first item packed representing minimum power that an appliance require to operate, will have at least the same or larger relative value than any item of size ΔP generated by the same appliance. Item with the same relative value are sorted in a way that, for items corresponding to the same appliance, we first pack the item corresponding to operating the appliance when applicable. The maximum additive gap with the optimal allocation for one time slot can be represented by the value obtained if the first item being rejected would have been scheduled. In Figure 4.4a, the bound corresponds to the total utility (i.e., total area of items) if item 4 is packed. Indeed, even if a smaller item can be packed such that the total weight is lower than the capacity C, total utility will always be smaller than that of packing the four items represented in the figure because of the ordering based on efficiency. Suppose that we only have these four items. Greedy algorithms will schedule items 1, 2 and 3. We can see that the optimal solution will be to schedule items 1, 3 and 4. These solutions are illustrated in Figure 4.4b. In this case, we can see the



(a) All items

(b) Scheduling solutions

Figure 4.4 – Greedy heuristics absolute performance bound for interactive loads

performance gap between greedy solution and the optimal one that is given by the are difference between each solution. This gap is smaller than the difference between the area of all four items and the area of items scheduled by the greedy algorithm. We note that this bound is valid for interactive loads because utility functions are independent between time slots. The case of time dependence is treated next when we analyze background loads.

background loads

For background loads, utility functions are dependent between time slots. Indeed, decision at a certain time slot affects utility that can be perceived at future time slots. Let us consider the example of batteries and pools for which a certain minimum operation duration is needed. Suppose that k time slots are needed for operation. We denote by $U(k)$ the utility if the appliance is operated over all k time slots and by $U(1)$ the utility of operating the appliance on one time slot. For this example, we can suppose that $U(k) > kU(1)$. This means that it is more attractive to provide power over needed operation duration than on a shorter time period.

On the contrary, for some appliances, operation at a certain time slot may render powering it on at future time slots less attractive. This is the case of heating systems that we take as a case study. Indeed, if heating is activated, it will raise temperature which makes heating less needed in subsequent time slots. This property of the utility function is called *sub-modularity*. Suppose S the set of items already scheduled and i an item that need to be packed. Global utility function U is sub-modular if it follows: $U(S \cup i) \leq U(S) + U(i)$.

An example that illustrates this inequality can be given in case of a heating system supposing two levels of utility named vital and comfort (see Section 4.2). Actually, adding an item for a certain heater (i.e. heating at a time slot) can fulfill vital needs which can transform later into comfort enhancement when other items are added to increase power allocation of the same heater.

It has been shown that, for greedy algorithms like the ones we propose, a multiplicative bound of $(1-\frac{1}{e}) \approx 63\%$ can be guarantied when objective functions are sub-modular (i.e., $UG \geq (1-\frac{1}{e})X^*$ [100, p.48]). In practice, better performance are expected from this strategy as we will see next.

Program-based loads

Due to the high dependence between time slots, the scheduling of program-based loads is equivalent to solving a multi-knapsack problem. We do not have performance guaranties of the proposed greedy algorithms for this case. However, a greedy approach based on utility efficiency of power allocation is expected to have good performance.

7 Numerical analysis

We now evaluate the performance of our schemes by instantiating them based on the case study presented in Section 4. The considered schemes are listed in Table 4.2. We begin by presenting the settings used in the analysis. Then, we present numerical results based on the supposed settings.

7.1 Settings

For the settings, we start by presenting system assumptions. Then, we present the parameters we use to instantiate our model. Finally, two use cases are introduced that allow to analyze schemes based on different utility function models for appliances.

7.1.1 System Assumptions

We focus on the functional groups introduced in Section 3. We remind/introduce the following assumptions:

- The system has no storage capabilities (neither at DSO/aggregator infrastructure nor at user premises). Indeed, energy produced at time t must be spent at time t (not $t' > t$) otherwise it is wasted.
- Power losses introduced by the distribution network are considered similar for all the users and hence the location of a user doesn't affect the decisions related to his power share. In particular, the net available capacity can be deduced straightforwardly from generation capabilities.
- The decision element DE_a knows the net available capacity as a function of time.
- Required exogenous information is known by both DE_h and DE_a (e.g., temperature forecast).
- The state of the appliances that are part of the system, their electric characteristics and the values of relevant variables (e.g. temperature at a home) at the beginning of the optimization period are known by DE_h and DE_a or by both. This information can be self-discovered by DE_h thanks to the maturity of technologies enabling autonomic capabilities.
- Time is slotted and the power consumed by an appliance is supposed to be constant in a given time slot. This assumption is mainly introduced to facilitate the discretization of the problem.

7.1.2 Parameters

In this section we fix several system parameters and exogenous variables and we focus on the behavior of the presented control schemes as a function of the available capacity. We fix the following general parameters:

- We select a slot duration of 5 minutes (τ).
- We suppose constant preference coefficients during the whole period: $\pi_v^a(h,t) = \pi_c^a(h,t) = 1\ \forall h\ \forall t\ \forall a$ (see Section 4).

- Temperature in homes evolves according to the simplified conductance/capacity model exposed in Section 4.2 and Section 6.2.2. We remind the dynamics:
$T_{ht} = T_{h(t-1)} + F(h)X_{ht}^2 + G(h)(T_e(t) - T_{h(t-1)})$.

- The DR period is set to $t_M = 100$ slots (≈ 8 hours).

- The system is made of $H = 100$ houses.

- We consider two types of appliances ($A = 2$): lighting (index $a = 1$) and heating (index $a = 2$)[1].

- Minimum tolerable temperature $T_m(h) := 15°C$ and preferred temperature $T_P(h) := 22°C$.

- We suppose a constant external temperature $T_e(t) = 10°C \ \forall t$ and an initial temperature $T_0(h) = 22°C \ \forall h$.

We suppose that the total available power is constant over the DR period, $C(t) = C$. We analyze the model for different values of C, ranging from low to full capacity (all appliances can be used). While this model is rather simple (two types of appliances, constant values for C, π and T_e), we believe that it is sufficient to give insights into requirements for each supposed control scheme.

The numerical analysis of the various presented MILPs (GM, RT, FGM and FRT defined in Section 5) has been carried out using IBM ILOG CPLEX v11.2 ([45]). As we will present next, depending on the use case, we may change CPLEX's parameter that sets the tolerable gap between the best integer solution found and the best possibly achievable total utility (LP relaxation). Since some problems take a long time to render a solution that satisfies desired gap, we set an additional stopping conditions that limits time of a resolution to 7 hours. Greedy algorithms are implemented using C++. All the schemes are run on a Linux-CentOS server with 2x4 core CPU, 16 GB of RAM and 2.66 Ghz clock speed.

7.1.3 Use cases

We differentiate two use cases based on utility functions of controlled appliances. In the first use case, only the comfort component of the utility functions is supposed. In the second use case, vital and comfort utilities are defined for each of the controlled appliances. Utility functions definition is based on the model presented in Section 4.

Use case 1

To study the importance of utility functions choice and its impact on the performance of the control scheme, we suppose one level of utility function per appliance (i.e., one level of criticality).

In this use case, this single utility model per appliance follows comfort component of utility functions defined in Section 4. This implies the following functions for each supposed appliance:

- Light utility linearly grows from $P_m^1(h)$ to $P_M^1(h)$ (see Figures 3.3b).

- For heating, utility linearly grows from $T_m(h)$ to the preferred temperature $T_P(h) := 22°C$ (see Figures 3.4b).

This setting does not mean that users have no vital needs, but that the decision entity considers that comfort utility curve is sufficient to give a "well enough" visibility into users' utility.

[1] Program-based appliances such as washing machines will be considered in the next chapters.

Use case 2
In this case, we suppose two levels of utility per appliance. This allows to better deal with fairness between appliances and between homes as opposed to only guarantying a worst case scenario for utility between homes (i.e., like it's done with FGM). It also allows to define different levels of service for each appliance.

To study the performance of the control schemes for several values of available capacity, we choose the following vital and comfort utility functions that are defined as follows:

- Vital utility for light is fully obtained as soon as the minimal light power $P_m^1(h)$ is reached, while comfort utility linearly grows from $P_m^1(h)$ to $P_M^1(h)$ (see Figure 3.3).
- For heating, vital utility linearly grows until the minimum tolerable temperature $T_m(h) := 15°C$ is reached, while comfort utility linearly grows from $T_m(h)$ to the preferred temperature $T_P(h) := 22°C$ (see Figure 3.4).

To speed up computation time for this use case, we consider a single resolution. Indeed, to take into account the two level utility model, the optimization should be run first considering vital utility maximization. Then, the same optimization need to be resolved supposing a different objective function that maximizes comfort utility while restraining to solutions that give the same total vital utility obtained as a solution of the vital utility maximization problem (see Section 4.4). Since, this requires to run twice similar optimization problems, we join the two objectives (maximize vital and comfort utility) to speed up the total resolution. So, we define a single objective function of the form $\max_{X_{ht}^a, x_{ht}^a} \sum_{t=1}^{t_M} \sum_{h=1}^{H} \sum_{a=1}^{A} \Gamma U_{v_{ht}}^{\ a} + U_{c_{ht}}^{\ a}$ in which the total vital utility is multiplied by a large constant Γ. This ensures the desired lexicographic ordering between solutions. In our numerical analysis, Γ is taken equal to the number of homes multiplied by 1000 ($\Gamma = 10^5$ in this case). Using the single objective formulation, the gap between the best solution found and the bound on possible improvement tolerated by CPLEX needs to be adjusted in order to ensure an acceptable solution quality for comfort utilities. Therefore, we lower the tolerated gap from its default value 10^4 and set it at $2.0/(\Gamma * 100)$.

7.1.4 Houses scenarios

For each of the considered use cases, we analyze proposed control schemes supposing homogeneous and heterogeneous houses. Parameters for each scenario are as follows.

Homogeneous houses
In this case, all homes have the same appliances characteristics and belong to the same class denoted by class 1. As shown in Table 4.1, we can define class 1 homes by the following minimum and maximum accepted power for each appliance (in Watts): Lighting system [50, 1000], Heater system [1000, 4000]. Thermal parameters $F(h)$ and $G(h)$ are taken equal to 0.0017 and 0.075 respectively. The parameter $F(h)$ means that if we heat at 1000W for 5 minutes (i.e., τ), temperature increases by $1.7°C$ (without any temperature loss). The value of the parameter $G(h)$ means that if there is a temperature difference of $10°C$ between indoor temperature and outdoor temperature, indoor temperature will decrease by $0.75°C$ over τ if no heating is provided. These values are chosen to stress the impact of heating over the optimization period.

Heterogeneous houses
In this case, homes are supposed to belong to two classes. Indeed, we consider that 50 homes belong to class 1 presented previously. The other 50 homes are of class 2 in which appliances have the following minimum and maximum accepted power (in Watts): Lighting system [50, 500], Heater system [1000, 2000]. As for thermal parameters $F(h)$ and $G(h)$, we suppose values equal to 0.0008

Class	$[P_m^1(h), P_M^1(h)]$	$[P_m^2(h), P_M^2(h)]$	$F(h)$	$G(h)$
1	[50, 1000]	[1000, 4000]	0.0017	0.075
2	[50, 500]	[1000, 2000]	0.0008	0.0365

Table 4.1 – Classes of houses

Short name	Full name
GM	Global maximum utility, ahead-of-time
RT	Maximum utility, real-time
FGM	Global maximum utility under fairness constraints, ahead-of-time
FRT	Maximum utility under fairness constraints, real-time
$GGreedy1$	Global greedy heuristic, ahead-of-time and exact utility
$RTGreedy1$	Real-time greedy heuristic and exact utility
$GGreedy2$	Global greedy heuristic, ahead-of-time and approximate utility
$RTGreedy2$	Real-time greedy heuristic, approximate utility

Table 4.2 – Centralized schemes reminder

and 0.0365 respectively. The value of $F(h)$ means that heating can increase the temperature by $0.8°C$ over τ if 1000W is provided to heating and no temperature loss occurs. This means that homes of class 2 are bigger than homes of class 1. The value of $G(h)$ means that without heating, indoor temperature decreases by $0.365°C$ over τ when temperature difference with the exterior temperature is equal to $10°C$. This means that homes of class 2 are better isolated than homes of class 1. Parameters defining Class 2 homes are summarized in Table 4.1.

7.2 Use case 1

Since it is based on direct control and the largest set of data, scheme GM provides the maximum total utility over the optimization horizon. We therefore take it as the benchmark to compare all the schemes. Figure 4.5 shows the results for the homogeneous case where all homes are of Class 1 (see Table 4.1).

Please observe that for facilitating the analysis, we will compare obtained total utility over the whole optimization period to the maximal feasible one. Hence, we normalize the total utility by dividing by the number of houses, the number of time slots and the number of appliances. This means that the absolute gain of the GM scheme compared with the others is larger than what is shown in the figures. Since utilities have been normalized based on maximum attainable utility, the maximum value that can be observed equals 1. Performance is expressed by the normalized total utility that we refer to by relative utility.

The extreme right points on the x-axis correspond to the power values at which, the total utility for most considered schemes reaches the maximum possible utility (i.e., represented by the value 1). Only the points in the graphs have been computed, the lines have been drawn to facilitate reading.

Figure 4.5 shows that in the homogeneous case, for this particular configuration, introducing fairness do not affect the relative utility (FGM and GM have the same performance) for most capacity values. Indeed, As discussed at the end of Section 5.3, even in a homogeneous setting, GM and FGM can have different relative utility outcome.

It is interesting to see, on Figure 4.5, that the real-time RT scheme significantly under performs scheme GM for most values of C. This is explained by the fact that, since the approach has no vision on the future, DE_a does not allocate enough power to the Heating system (it does not predict the utility that can be generated in the future). Therefore, for low values of C, even when it can allocate power for the Heating system of some houses, it prefers to allocate the whole power to the Lighting system which has a better utility efficiency (i.e., ratio of utility value and amount of power required). For values of C smaller than 100000W, RT can allocate power for heating but it does not. This is actually due to the utility function model considered that is more adapted for an optimization over a period of time. Indeed, when temperature drops below minimum tolerated temperature $T_m(h)$ set to $15°C$, if powering the appliance cannot bring the temperature above this value, heating is never scheduled subsequently. This shows that the real-time RT mechanism is not well adapted and that additional intelligence is needed when a real-time approach is required like the one introduced with greedy approximations. Indeed, we can see that greedy algorithms are capable of better accounting for the value of heating whether they are global or real-time. However, for high capacity values, heating value is overestimated which lowers the performance of these algorithms except for $RTGreedy2$. This can be seen in Figure 4.6a that illustrates average temperature for capacity $C = 1.5 \times 10^4$. We can see that $RTGreedy2$ maintains temperature above the preferred value of $22°C$ while other greedy algorithms tend to overheat. Indeed, based on the way $RTGreedy2$ is defined, the utility of heating is equal to zero when temperature is above preferred value. This can also be seen for low capacity as shown in Figure 4.7a for $C = 0.4 \times 10^4$. This figure also shows that RT and FRT do not provide any heating for this capacity value. Furthermore, FGM does not maintain temperature above minimum tolerable temperature ($15°C$). In Figures 4.6b and 4.7b, minimum temperature perceived at each time slot for the capacities $C = 1.5 \times 10^4$ and $C = 0.4 \times 10^4$ is represented respectively. For $C = 1.5 \times 10^4$, we can notice that even though the same average temperature is obtained with RT and FRT, RT stops allocating heating for some homes. For $C = 0.4 \times 10^4$, we can see that minimum temperature is higher for FGM.

As for the heterogeneous case where homes are of classes 1 and 2 (see Table 4.1), results are presented in Figure 4.8. In this figure, we can see that $GGreedy1$ has strategies that are close to GM in the sense that homes of class 2 are favored over homes of class 1. Indeed, this is not surprising since homes of class 2 have a better energy efficiency. So, in the absence of fairness, it is best to allocate heating to class 2 homes first. We can notice that, as expected, FGM provides similar relative utility to both classes of homes. Algorithms $GGreedy2$ and $RTGreedy2$ gives more priority to homes of class 1 since temperature gains from heating in time slots following that for which heating decision is taken, are underestimated. For all greedy algorithms, heating utility is overestimated compared to lighting. This is depicted by low performance for high values of available capacity.

On the overall, we can see that approximate heating utility computation for real-time schemes ($RTGreedy2$) provides a better performance than the exact computation ($RTGreedy1$).

To sum up, fairness provided by FGM and FRT may not be sufficient even for a homogeneous case. It cannot guaranty an equal minimum heating requirements for homes. This is due to the fact that maximizing the wost case scenario between homes do not necessarily capture requirements related to individual appliances use.

7.3 Use case 2

We now analyze the second type of fairness that can be guarantied by the utility function model and allows to guaranty minimum requirements between appliances and thus between homes.

To analyze the performance of the proposed control schemes, we also study a homogeneous scenario where all homes belong to Class 1 and a heterogeneous scenario where the total number of homes is equally divided among Classes 1 and 2 (see Table 4.1).

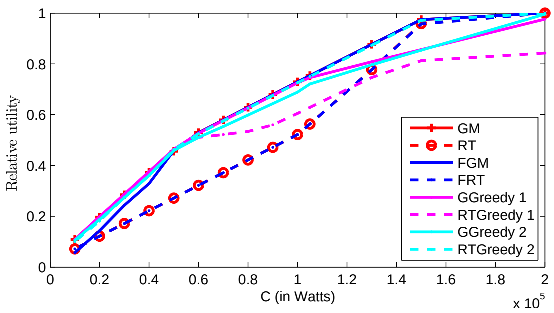


Figure 4.5 – Relative utility per home over the DR period as a function of the available capacity for centralized schemes (Single utility model, homogeneous case)

As in the previous use case, we are dealing with utilities that have been normalized. So, the maximum utility for an appliance is equal to 1 for both vital and comfort utility functions. Performance is again expressed by the relative utility provided to homes over the whole optimization period. Only the points in the graphs have been computed.

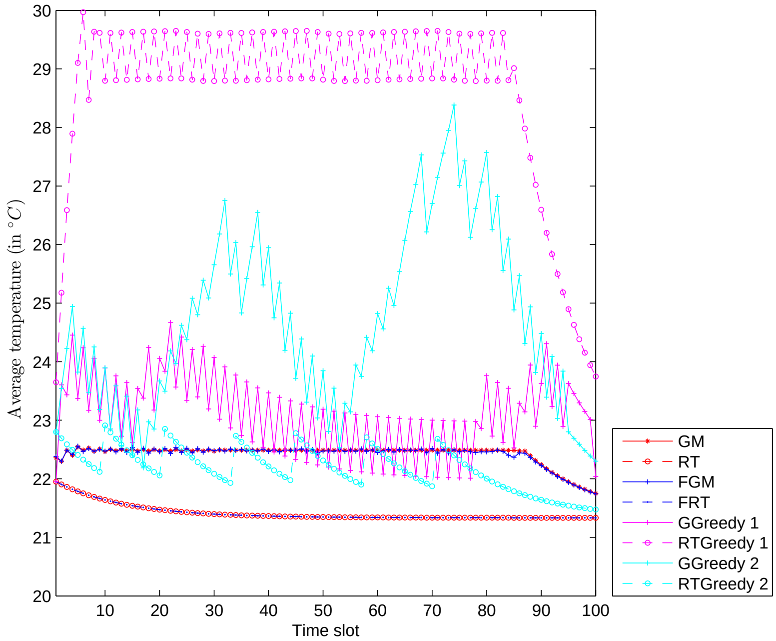
To compare performance, we remind the lexicographic order between vital utility and comfort one. So when comparing two points (corresponding to two schemes) for a certain value of capacity C, we first compare the vital values before comparing the comfort ones.

We can see on Figure 4.9 that both ahead-of-time schemes GM and FGM have very similar performance. This remains valid even when taking the heterogeneous case. This shows that having a multi-level utility model may be sufficient to enforce fairness. In addition, as discussed in Section 3.1, the decision entity might be willing to provide different contracts depending on maximum utility that can be perceived by a certain home which makes guarantying the worst case scenario more advantageous for homes with a low contract level (i.e., a low total utility value assigned).

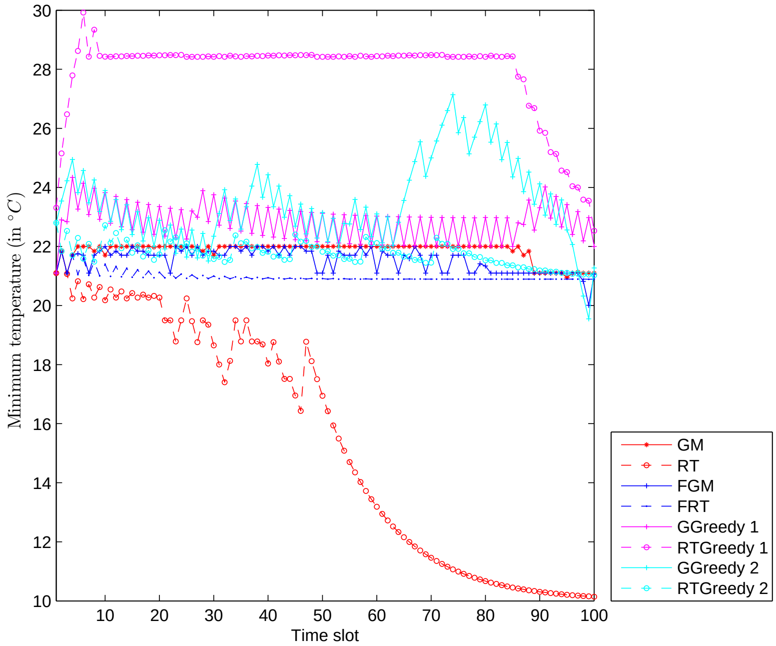
As for real-time schemes RT and FRT, we can see in Figure 4.9a that they allow to guaranty a maximum vital utility. However, they cannot reach the optimal performance for comfort enhancement (see Figure 4.9b). This is mainly due to the fact that lighting is always prioritized over heating since visibility on the real value of heating over the DR period is not considered by those schemes.

For greedy algorithms, we see that they are very close to GM for most capacity values. Indeed, $RTGreedy2$ is always optimal. In Figure 4.9a, we can see that $GGreedy2$ does not provide maximum relative utility. Indeed, when decision can be taken on any slot, heating is allocated for slots where a vital need is expressed depending on packing opportunities. For a homogeneous case, some homes will have heating at time slots where temperature is lower than minimum tolerated temperature $T_m(h)$. Since capacity is limited, other homes will be allocated the minimum power for heating in later time slots. However, since packing decisions are taken for any time slot, this may create situations in which all available capacity is used for some homes so that their temperature is increased significantly (by adding blocks of ΔP=50 Watts) and no capacity is available to give other homes their minimum required power for heating. These situations are not likely to happen with the real-time greedy algorithms since decision usually favors the most disadvantaged homes.

For high capacities, $GGreedy1$, $GGreedy2$ and $RTGreedy1$ have low comfort performance as seen in Figure 4.9b. Indeed, this is due to the overestimation of heating utility due to the sub-modular property of heating as discussed for the previous use case. This is not the case for $RTGreedy2$ since no incentives are given to have temperature above the preferred value $T_P(h)$. We can notice that GM do not show the maximum utility for capacity $C = 0.5 \times 10^4$. This is due to the complexity of finding the optimal solution using CPLEX that is stopped if no optimal solution is provided in reasonable time. This shows the high value of simple heuristics capable of reaching good performance.

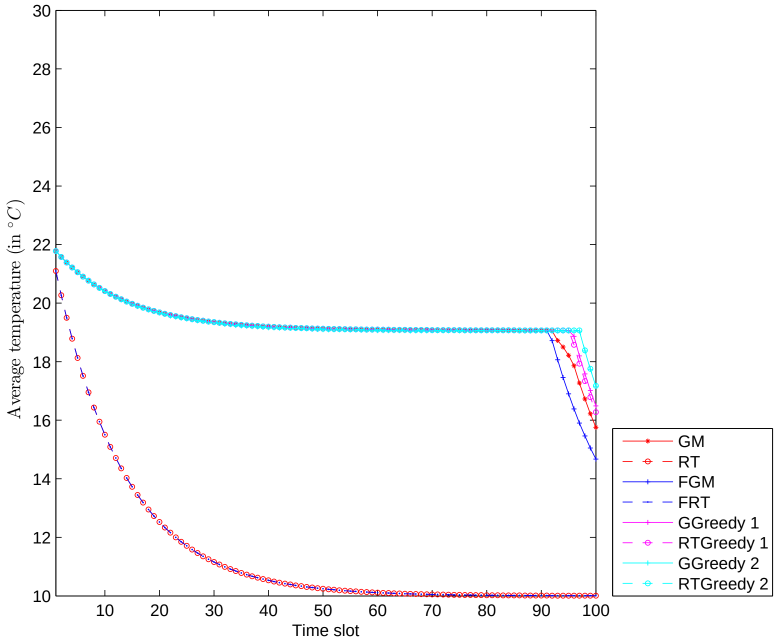


(a) Average temperature

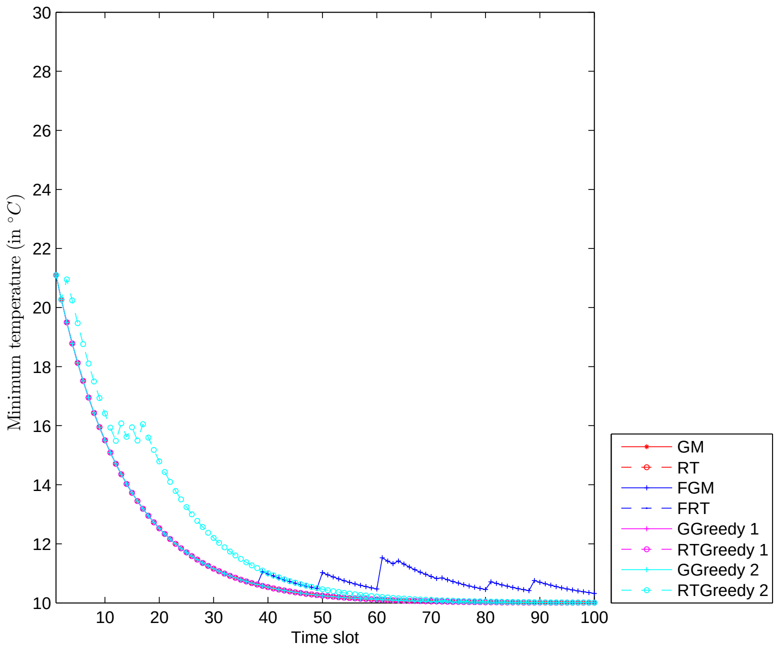


(b) Minimum temperature

Figure 4.6 – Temperature evolution over the DR period for $C = 1.5 \times 10^4$W by centralized schemes (Single utility model, homogeneous case)

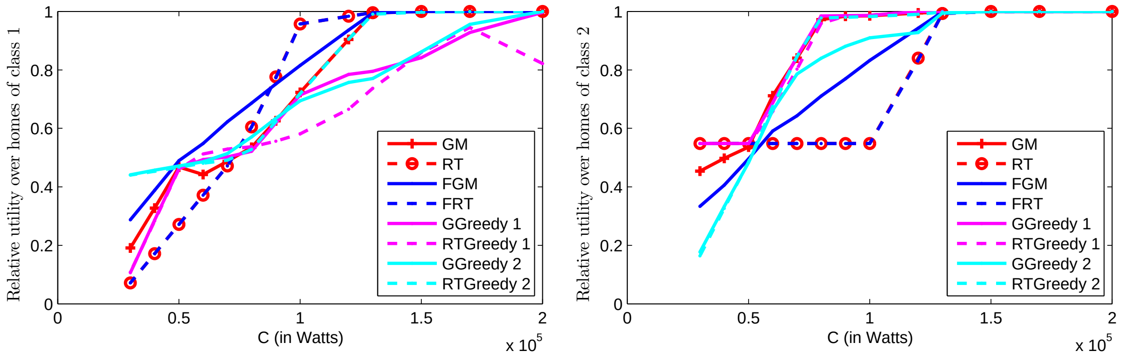


(a) Average temperature



(b) Minimum temperature

Figure 4.7 – Temperature evolution over the DR period for $C = 0.4 \times 10^4$W by centralized schemes (Single utility model, homogeneous case)



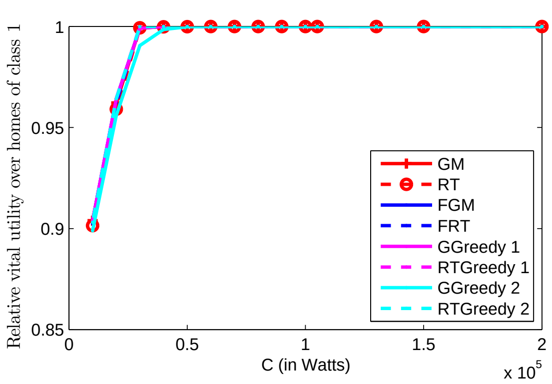
(a) Homes of class 1 (b) Homes of class 2

Figure 4.8 – Relative utility per home over the DR period as a function of the available capacity for centralized schemes (Single utility model, heterogeneous case)

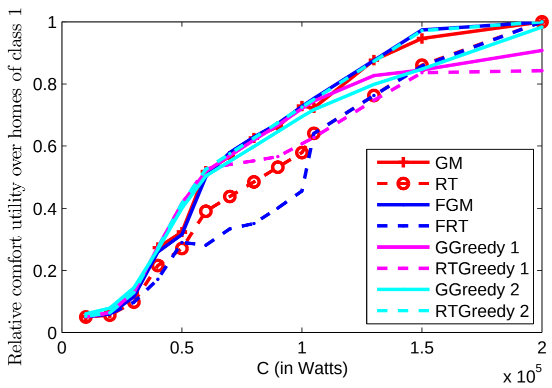
Let us now analyze temperature evolution for capacities $C = 1.5 \times 10^4$ and $C = 0.4 \times 10^4$ to compare with the previous use case. Figures 4.10 and 4.11 show average and minimum temperature for each capacity value. We can see that thanks to the vital utility, RT and FRT maintain temperature above minimum tolerable temperature ($15°C$) for all homes. This is also the case for all other schemes except for $GGreedy2$ (for the reasons previously discussed).

For the heterogeneous case, results are shown in Figure 4.12. Schemes RT and FRT struggle to reach the maximum vital performance for low capacities in which heating cannot be provided to all homes simultaneously. Performance for comfort enhancement in the heterogeneous case are similar to the homogeneous one.

For greedy algorithms, comfort utility is usually poor for high available capacity values especially for $RTGreedy1$. Indeed, since $RTGreedy1$ looks at temperature evolution in subsequent time slots to take decision at a time slot t. It can see that temperature will go below $T_m(h)$ (with the assumption that no heating is provided subsequently). This gives a value of heating at a slot t that have a vital component even if vital needs are already fulfilled at t and if later the scheduler can provide additional heating that prevents temperature from going below $T_m(h)$. This is also why $GTGreedy1$ and $RTGreedy1$ favors homes of class 1.

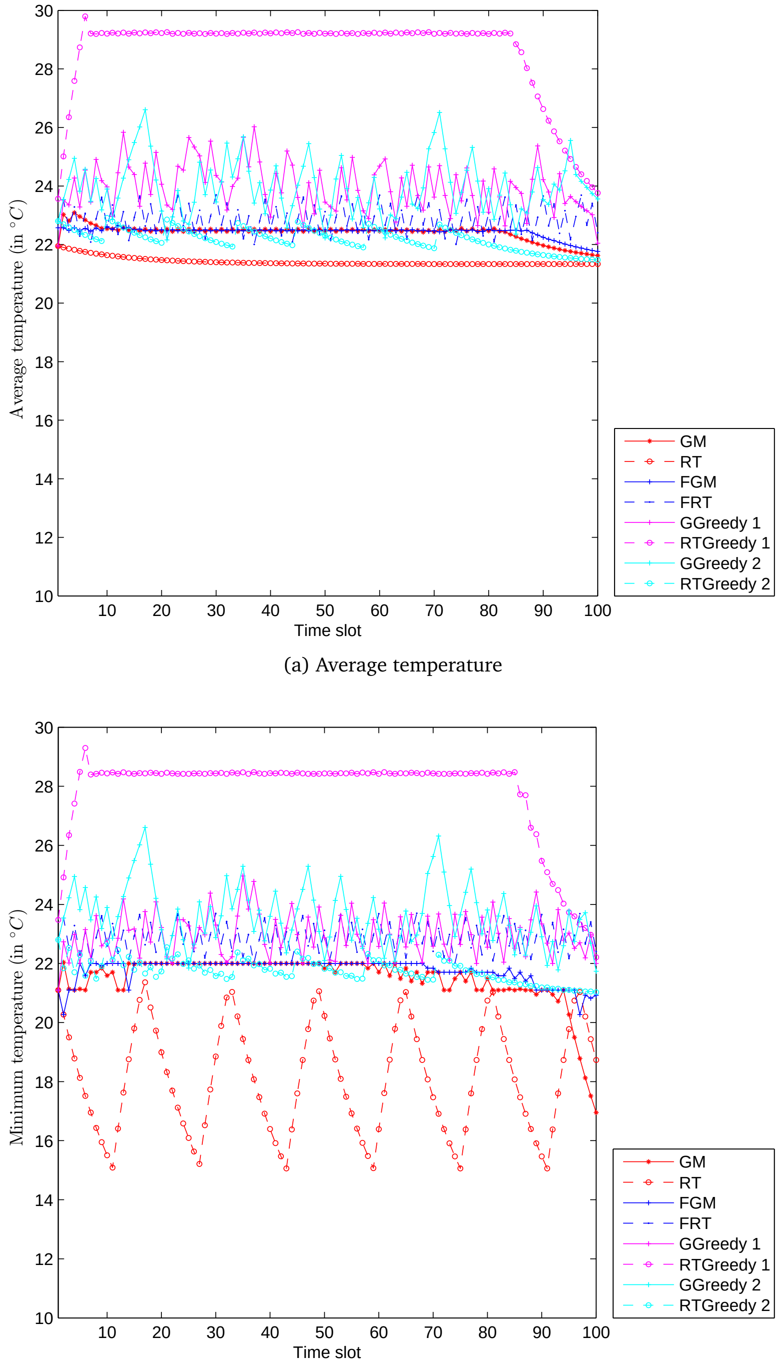


(a) Vital utility



(b) Comfort utility

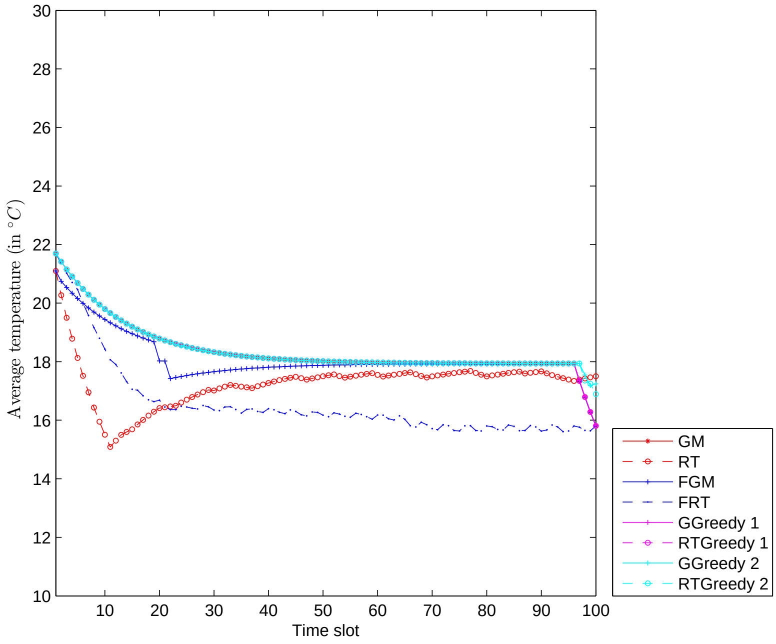
Figure 4.9 – Relative utility per home as a function of the available capacity for centralized schemes (Double utility model, homogeneous case)



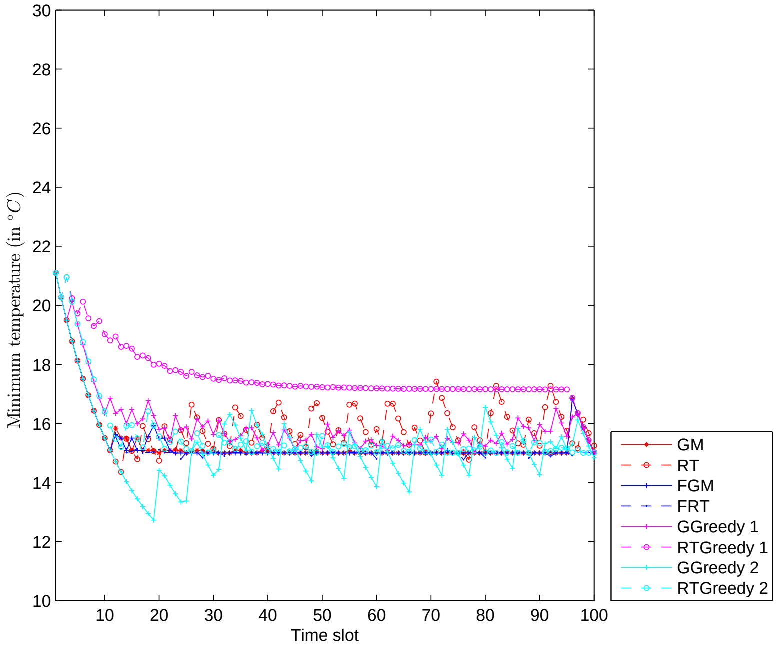
(a) Average temperature

(b) Minimum temperature

Figure 4.10 – Temperature evolution over the DR period for $C = 1.5 \times 10^4$W by centralized schemes (Double utility model, homogeneous case)

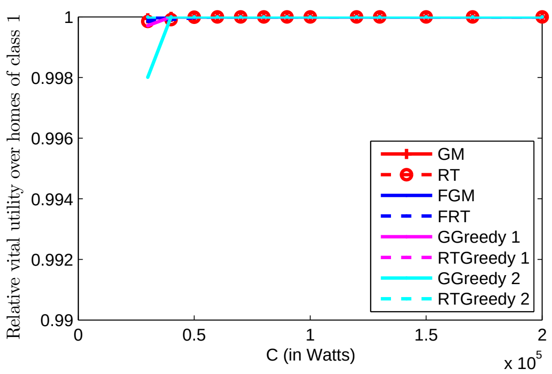


(a) Average temperature

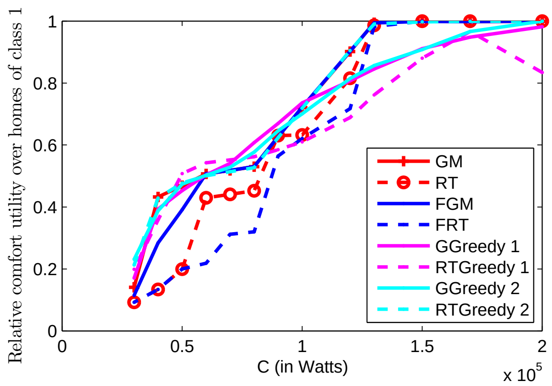


(b) Minimum temperature

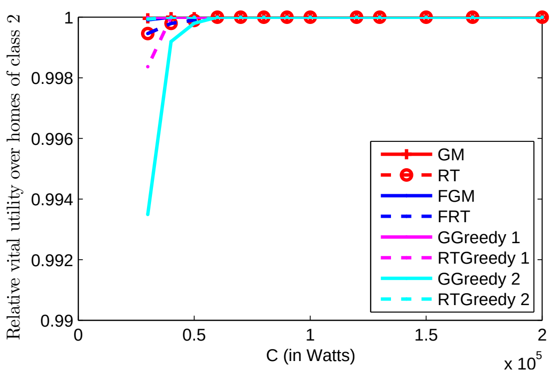
Figure 4.11 – Temperature evolution over the DR period for $C = 0.4 \times 10^4$W by centralized schemes (Double utility model, homogeneous case)



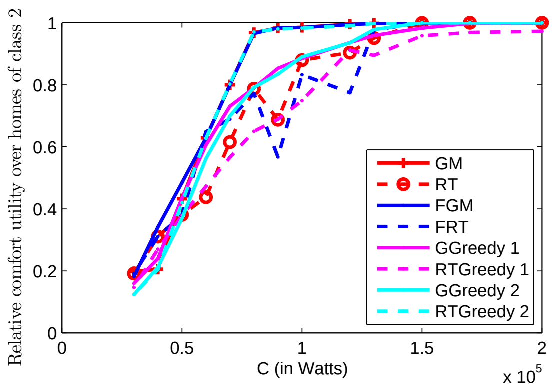
(a) Vital utility of class 1 homes



(b) Comfort utility of class 1 homes



(c) Vital utility of class 2 homes



(d) Comfort utility of class 2 homes

Figure 4.12 – Relative utility per home as a function of the available capacity for centralized schemes (Double utility model, heterogeneous case)

8 Conclusion

In this chapter, we presented DR solutions based on a centralized architecture. Theses solutions are capable of controlling individual appliances power consumption at homes while aiming at maximizing global utility perceived by users under system constraints (e.g. available capacity, fairness constraints). The utility functions used to produce control decisions rely on users' setup and on measurement of relevant homes' variables. We show that schemes can be designed in view of the time horizon of the control, the decision timescale (ahead-of-time or real-time) and the algorithm used to find power allocations. We considered exact resolution (MILPs) and approximate resolution (greedy heuristics).

Through numerical analysis, we showed the importance of an adapted design to capture the real value of appliances. This can be achieved when the period of DR and system parameters are known ahead-of-time. We observed that all greedy algorithms perform well while significantly improving scalability. We also observed that an exact computation of the real value of operating appliances may not be necessary. This was shown by the near-optimal performance of the greedy algorithm that acts in real-time while approximating appliances value. Finally, we showed that considering multi-level utility model that allow to deal with needs' criticality is sufficient to introduce fairness between homes by enforcing it between appliances.

Centralized DR solutions provide DSOs and aggregators the capability to evaluate the impact of control on users' utility. The possibility to take into consideration the quality of experience of users enables new services and business models.

Chapter 5

Partially distributed control

In the previous chapter, we saw that a centralized solution based on full, fine-grained, available information can be used to address the problem of scheduling appliances given an undersized capacity budget. However, centralized solutions have the following limitations.

- Finding the optimal global solution can be costly in term of computational power (see Section 6.1.1). This can be mitigated by using heuristics (see Section 6.2), but the trade-off is a lower performance.
- All input variables and output decisions must be directed in a central node. This creates a communication overhead that may be non negligible and have some impact in term of responsiveness.
- Gathering complete information on users' appliances (characteristics, planned usage) is a huge privacy breach that can affect the acceptance of the control schemes by end users.
- The central node in charge of computing the solution is a potential single point of failure.

A natural way to address all the issues above is to reduce the importance of the central entity. In this chapter, we focus on mechanisms where the central entity only provides coarse-grained allocations and delegates part of the problem to lower levels. This corresponds to *vertical communication* schemes introduced in Section 1.2[1].

To fix ideas, we suppose like in previous chapter a two level architecture made of an aggregator decision entity DE_a at one level and home decision entities DE_h at the other level. The DE_a needs to cap consumption during a certain known period of time. To do so, it sends control signals to homes' DE_h. Based on these signals, each DE_h schedules appliances in the way that is the most satisfactory in terms of utility for users.

In more details, one can envision two categories of DR mechanisms based on the information used by the DE_a.

One-way hierarchical schemes Inputs for decision-making at higher level is limited to static informations that are collected in advance (e.g. at the time of setting the contract for the service). In this case, information exchange between entities is top-bottom only. Figure 5.1 illustrates that type of scheme.

[1]Horizontal communications, where lower level entities collaborate to improve their situation with minimal or no intervention from higher entities, will be considered in the next chapter

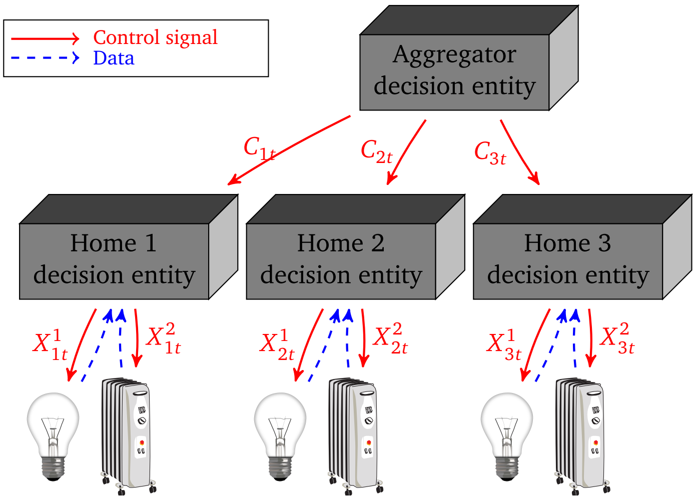


Figure 5.1 – Local control

Two-way hierarchical schemes In addition to static information, the central entity also collects a limited feedback from lower levels to adjust its coarse-grained allocation. This is illustrated by Figure 5.2.

The main contribution of this chapter is the introduction of a new two-way hierarchical scheme based on a subgradient method. We evaluate its performance against the centralized solution from previous chapter and two canonical one-way hierarchical schemes in several use cases that mix interactive, background and program-based appliances.

The chapter is organized as follows: Section 1 presents related work. Section 2 introduces proposed schemes that are separated into mechanisms that only rely on one-way communication from the aggregator to homes and a scheme that suppose two-way communication between the aggregator and homes. To analyze the performance of proposed schemes, several use cases are considered and introduced in Section 3. In Section 4, numerical results are presented for each use case. Finally, Section 5 presents general conclusion with respect to supposed schemes.

Most of the content presented in the present chapter has been published in [KKP14, KKMP15], in collaboration with Daniel Kofman, Fabien Mathieu and Michal Pioro.

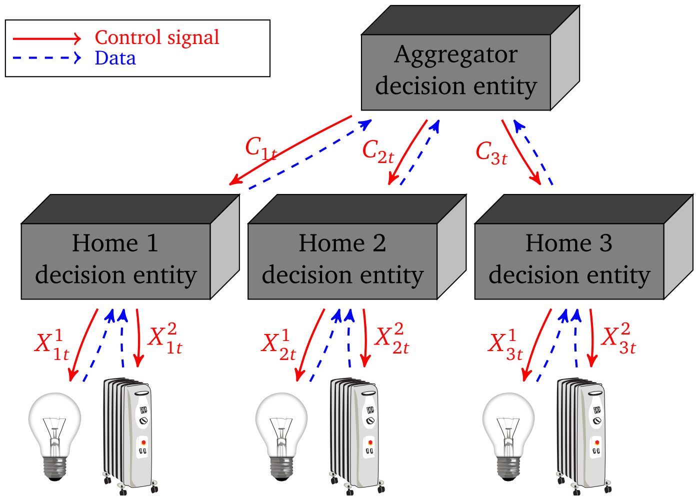


Figure 5.2 – Local control with feedback

1 Related work

The idea of using vertically distributed control schemes with no or limited feedback is quite natural. Indeed, the literature contains a significant amount of proposals based on hierarchical control schemes (with feedback) in the context of limiting consumption capacity to a certain desired value (e.g., [30, 97, 55, 89]).

However, these proposals usually address the problem by taking the dual of system capacity constraint. These dual variables can be seen as prices for time slots (e.g., [30, 97, 55]). For this reason, these schemes are usually presented as DR schemes based on pricing, not direct control[2].

The proposals in [97, 55, 89] are examples of schemes that are designed for residential consumers and that can take into account flexibility of generic appliances. Authors in [97] propose a customer reward scheme that encourages users to accept direct control of loads. They propose a time-greedy algorithm (maximizes utility slot by slot) based on the utility that each appliance declares for each slot. As discussed in the previous chapter, instantaneous value of an appliance have to be carefully evaluated to capture the real benefit from using this appliance (e.g., heating system).

Authors in [55] propose a dynamic pricing scheme based on a distributed algorithm to compute optimal prices and demand schedules.

The most closely related work to our proposal is the direct control scheme presented in [89] which is very similar to [55] if prices are interpreted as control signals. The authors in [89] propose to solve a problem similar to ours, but in their approach, intermediate solutions can violate the constraints so that convergence of the algorithm is required (like all other schemes based on dual decomposition) to produce a feasible allocation that satisfies total capacity constraint. The authors do not discuss scalability and communication requirements in terms of the number of iterations required. They also assume concave utility functions. Moreover, the proposed scheme still requires disclosure of extensive information to the central entity (i.e., home consumption profile), so the approach is not adapted to reduce privacy issues.

In the present work, we target to better deal with privacy while guaranteeing the fulfillment of capacity constraints even during intermediate computation.

[2]Actually, they can also apply as direct control mechanisms where prices represent control signals, but the resulting scheme does offer the same guarantees that we try to achieve in the present work.

2 Proposed schemes

In this section, we first present two simple one-way schemes (no feedback). Such schemes have the advantage of being free of privacy and scalability issues. They will be used, along with the optimal centralized solution from the previous chapter, to assess the performance of the two-way scheme we present in the second half of this section.

2.1 One-way hierarchical schemes

Since only static home information is available to DE_a in a one-way scheme (no fine-grained nor coarse-grained dynamic information), its decision cannot be directly related to improving users satisfaction. Instead, we focus on schemes in which a certain level of fairness is introduced. This allows us to improve users' acceptance. We now propose two one-way schemes that differ in the time repartition of the available capacity.

2.1.1 Local maximum utility

In this scheme, homes are grouped into classes. Each class receives a certain constant amount of power based on parameters associated to the class. To fix ideas, we assume that the allocation is done based on the maximum power controlled appliances consumed in homes of the class in nominal conditions.

This maximum power is referred to by subscribed power and is denoted by $L(h)$ for each home h. Homes in the same class will have the same value for subscribed power.

Suppose $C(t)$ the targeted capacity limit by DE_a for time instant t. At each decision time, DE_a allocates power to homes proportionally to their subscribed power. This allows to introduce fairness in power allocation based on the share (in power consumption) of the home being controlled. So, the power limit allocated to home h for time slot t is:

$$C_{ht} = \frac{L(h)}{\sum_i L(i)} C(t).$$

Based on the received limit capacity C_{ht}, DE_h at each home h decides the corresponding allocation per appliance by solving the restriction of the centralized problem (4.1) to h, using C_{ht} instead of $C(t)$. This control scheme is named *Local Maximum* utility and is denoted by LM.

2.1.2 Rolling Blackout

In this scheme, the power allocated to each home by DE_a depends on the time and is either the maximum subscribed power $L(h)$ or zero. Fairness in power allocation is introduced by implementing a round-robin like procedure. We introduced this scheme because it is very close to the way capacity shortage is handled for power grid systems [37]. Indeed, when planned blackouts are scheduled, a similar approach is used for connecting and disconnecting districts. The main difference compared to the present scheme is the granularity of decisions (districts versus homes).

In details, having a list of homes, we start by allocating maximum subscribed power to the maximum number of users following an arbitrary circular list order[3]. After a while, these users are disconnected and the available power is provided to the next users in the list. When the list is circular so the scheme can go on indefinitely. The duration of power delivery is decided by DE_a and is denoted by t_{RB}. This period represents the minimum number of time slots homes are connected before being disconnected.

[3] Remaining capacity, if any, is allocated to the next user in line. That user is still considered disconnected for the scheduling.

To summarize, for each home h, $C_{ht} = L(h)$ for t in which homes are connected and $C_{ht} = 0$ otherwise, with the possible exception of one left-over allocation.

Like in the previous approach, DE_h at each home h decides the power allocation per appliance by solving the restriction of problem (4.1) to h, using C_{ht} instead of $C(t)$

We call this scheme Rolling Blackout RB.

2.2 Greedient approach

We now propose a two-way scheme that aims at achieving a trade-off between a centralized solution that provides maximum performance in terms of total utility value, and local one-way schemes that enforce scalability and privacy.

To reach privacy and scalability goals with limited feedback, we propose a simple primal decomposition of the global GM problem into a master problem, described in (5.1), and subproblems, described in (5.2).

Master problem

$$\max \sum_{h=1}^{H} U_h \tag{5.1a}$$

$$\sum_{h=1}^{H} C_{ht} = C(t), \;\; \forall \, t \tag{5.1b}$$

$$C_{ht} \geq 0, \;\; \forall \, h \;\; \forall \, t. \tag{5.1c}$$

Subproblems

For each home h, the following Mixed Integer Linear Problem is solved:

$$U_h = \max \sum_{t=1}^{t_M} \sum_{a=1}^{A} U_{ht}^{a} \tag{5.2a}$$

$$\sum_{a=1}^{A} X_{ht}^{a} \leq C_{ht}, \;\; \forall \, t. \tag{5.2b}$$

If the C_{ht} are known, the subproblems (5.2) are no different from one-way schemes and can be solved accordingly. The main issue to address is the handling of the master problem (5.1): how to shape a per-home allocation of capacity that maximizes the utility, given that one would like the full characteristics of appliances to remain private?

To treat this problem, we propose a new heuristic called the Sub-Greedient method (SG). This heuristic is inspired by the Sub-Gradient method [12], but is adapted to take into account the specificities of our model. In particular, we introduce the notion of *Greedient*[4], inspired by the greedy algorithms from Section 6.2. Greedients will be used instead of more traditional (sub)-gradient approaches to estimate the utility meso-slope of a given house.

We briefly describe the main steps of SG:

- SG needs to be bootstrapped with an initial power allocation.
- DE_a transmits to each home DE_h the current allocation proposal C_{ht} $\forall t$. DE_h then solves the corresponding subproblem (5.2) like for a one-way scheme. It sends back the total utility U_h feasible, along with the Greedient associated to the current solution.
- Using the values reported by homes, DE_a then tries to propose a better solution.
- The process iterates for up to K_{MAX} iterations. In the end, the best solution found is used.

We now give the additional details necessary to have a full view of the solution.

[4] We introduce that term to remind of the gradient and because it is very similar in spirit to the metric used to evaluate allocation items in the Greedy algorithms. The term *discrete gradient* could have been used instead, although the greedient, which will be formally defined below, differs from the usual definition of a discrete gradient [5]

2.2.1 Initial allocation

The first allocation (before the first feedback) is necessarily a one-way allocation. We propose to use the *RB* schemes described above. The interest for SG of such an initial allocation (e.g. compared LM) is that it breaks possible symmetries between homes and gives an initial diversity that will help finding good Greedients.

2.2.2 Greedient

We define the greedient g_{ht} as the best possible ratio between utility improvement of home h and capacity improvement at time t. Formally, if $U_h'(\Delta C_t)$ represents the best feasible utility for home h if its current allocation is increased by ΔC_t at time t, we have

$$g_{ht} := \max_{\Delta C_t > 0} \frac{U_h'(\Delta C_t) - U_h}{\Delta C_t}.$$

To compute the greedient of a home, we define the greedient g_{ht}^a of an appliance a as follows: for a given allocation C_{ht}, $C_{ht}^0 \geq 0$ represent the capacity unused by house h at time t in the optimal allocation. $U_h^{a\prime}(\Delta C_t)$ represents the maximum utility for appliance a if an additional capacity of up to ΔC_t is added its current consumption. Then we have

$$g_{ht}^a := \max_{\Delta C_t > 0} \frac{U_h^{a\prime}(C_{ht}^0 + \Delta C_t) - U_h^a}{\Delta C_t}.$$

It is easy to see that the greedient of a home is greedient of its best appliance : $g_{ht} = \max_a g_{ht}^a$.

Note that if we suppose that the utility functions have a diminishing return property, which is the case for the use cases we considered, the greedient of an appliance is equivalent to the gradient of the utility function in the situations where $C_{ht}^0 = 0$ and continuous variation of power is allowed: for these situations, the best efficiency is observed for $\Delta C_t \to 0$. The only difference (under diminishing return assumption) is when there are gaps between allowed allocations: the greedient will point to the next allowed value while the gradient will report 0.

Remark The improvement promised by the greedient is only valid for a specific step (or range) of capacity increase. However, this step is not disclosed to DE_a to prevent the central entity to infer the characteristics of users based on their inputs. As a result, the greedient hints at the potential interest of investing additional capacity to a given home, but it is not reliable. This is the price we choose to pay to limit privacy issues.

2.2.3 Finding better solutions

To update the current solution at the k-th iteration, DE_a does the following:

- It first computes values $\alpha_k g_{ht}$ $\forall h \forall t$. These values represent potential increase of C_{ht}. Parameter α_k, called the *step size*, which will be described below.
- It then adjusts the new values of C_{ht} so they stay positive and fit the capacity constraints.

For the adjustment phase, it is important to deal with cases where allocation update $\alpha_k g_{ht}$ is larger than available capacity $C(t)$ or even maximum subscribed power $L(h)$ of home h, so we first cap $\alpha_k g_{ht}$ at the minimum between power limit of the smallest home ($L_m := \min_h L(h)$)[5] and system capacity $C(t)$. We therefore define $\beta_{kht} = \min(\alpha_k g_{ht}, Lm, C(t))$.

[5] We chose the capacity of the smallest home instead of the capacity of the current home to avoid a *masking* effect where the demands of larger homes cloud the demands of smaller homes.

Then for each t, we remove some positive common value λ_t to the C_{ht} to keep the sum of the allocations equal to the total capacity $C(t)$. To avoid houses with low C_{ht} to be badly impacted (in particular to avoid negative allocations that will be impossible to enforce), a subset I_t of the houses will be "protected" so that their values cannot decrease. In details, we do the following, starting with $I_t = \emptyset$:

- We compute λ_t such that the values

$$C'_{ht} = \begin{cases} C_{ht} + \max\{\beta_{kht} - \lambda_t, 0\} \text{ if } h \in I_t, \\ C_{ht} + \beta_{kht} - \lambda_t \text{ otherwise,} \end{cases} \tag{5.3}$$

  sum to $C(t)$. See [79, 40] for more details.

- We protect (e.g. add to I_t) all houses that get a negative value C'_{ht}.

- We iterate the steps above until all C'_{ht} from eq. (5.3) are positive. DE_a then proposes C'_{ht} as a new solution to investigate.

Remarks While the solution described here applies to a 2-level hierarchy (DE_a, DE_h), it can be generalized to M levels to take into account static maximum capacity of different aggregation points on a hierarchical distribution network: considering an aggregation point m at a certain level, the greedient for m is the maximal greedient of its children.

The adjustment phase can take into account capacity constraints of m, such as static power limits at each level of the hierarchical distribution network.

Also note that the proposed scheme can run asynchronously in the sense that it does not require all houses to communicate simultaneously. In fact, as soon as at least two homes respond, a local reallocation can be made for responding homes without having to wait for others to respond: we just need to restrict the problem to the corresponding subset of homes, using their current cumulated allocation as capacity limit.

2.2.4 Choosing the step size

The step size α_k for each iteration k is a crucial parameter. Indeed, choosing appropriate step sizes is key to speeding up resolution. Intuitively, large values of α_k should be chosen to make the allocation update (dictated by $\alpha_k g_{ht}$) useful for high consumption appliances, and lower values are more adapted to low consumption appliances.

Among the step size sequences proposed for subgradient methods, we consider for our performance analysis the two following ones (see [12]):

A diminishing non-summable step size rule of the form $\alpha_k = \frac{a_1}{\sqrt{k}}$.

A constant step length rule of the form $\alpha_k = \frac{a_2}{||g_{ht}||_2}$, where $||g_{ht}||_2$ is the euclidean norm of the vector of all greedients.

Please note that the second step size method (named following [12]) ensures a constant step length ($||C'_{ht} - C_{ht}||_2 = a_2$) when no capacity constraints are considered.

The value of parameter a_1 (resp. a_2) is also critical for the performance of the scheme. It currently manually adjusted to provide the best result, but we believe that an automatic estimation of the best value given the static parameters of a given use case is a promising lead for future work.

3 Considered Use cases

We propose to evaluate the performance of proposed solutions in use cases very similar to the one from the previous chapter. There are roughly two main differences:

- We focus on the case where utility is described by vital and comfort components (corresponding to the use case 2 of Section 7.1.3).
- We introduce an additional appliance representative of program-based loads: washing machine(s).

In more details, we consider a constant total available power over the DR period ($C(t) = C$). We analyze the model for different values of C ranging from scarce resources (impossible to completely fulfill all vital utilities) to reasonable resources (all needs can be fulfilled using a centralized allocation with full information). Indeed, these values correspond to different powering on capabilities (ranging from one appliance to all types of appliances supposed). While we consider a simple model where constant values for C, π and T_e are supposed, we believe that it provides sufficient insights to capture the trade-off between the efficiency of an allocation and the privacy of the users.

To study the performance of the control schemes for several values of capacity, we choose the following system parameters:

- The DR period is set to $t_M = 100$ slots (≈ 8 hours).
- The size of the system is $H = 100$ houses.
- We suppose a constant external temperature $T_e(t) = 10°C$ $\forall t$ and an initial temperature $T_0(h) = 22°C$ $\forall h$.

We consider two use cases that will differ in the appliances supposed. The first case is taken from previous chapter: only appliances for which decision can be made on a time slot without having to commit for a specific duration are supposed (e.g. lights and heaters). For the second case, an additional type is supposed corresponding to program-based loads: when the appliance is operated it must be fed until it completes its program.

Use Case: Lights and heaters
The first use case corresponds to the one described in Section 7.3. We consider two types of appliances ($A = 2$): lighting (index $a = 1$) and heating (index $a = 2$). Utility functions for all appliances have a vital and a comfort component. Vital light utility reaches its maximum value as soon as the minimal light power $P_m^1(h)$ is provided, while comfort utility linearly grows from $P_m^1(h)$ to $P_M^1(h)$ (see Figure 3.3). For heating, vital utility linearly grows until the minimum tolerable temperature $T_m(h) := 15°C$ is reached, while comfort utility linearly grows from $T_m(h)$ to the preferred temperature $T_P(h) := 22°C$ (see Figure 3.4).

Two types of houses are considered for this use case: Classes 1 and 2 (see Table 5.4). Class 2 homes have a better energetic performance than class 1 ones (less light power required to achieve full utility and better insulation). For the performance analysis, we will consider a homogeneous scenario with 100 houses of class 1, and a heterogeneous scenario with 50 houses of class 1 and 50 houses of class 2.

Use Case: Lights, heaters and washing machines
For this use case, we add washing machines (index a = 3) to the previously supposed appliances (A = 3). We suppose that it has a non-interruptible operation once it is turned on (see Section 4.3).

Type	$P_m^1(h)$	$P_M^1(h)$
1	50	1000
2	50	500

Table 5.1 – Lighting parameters

Type	$P_m^2(h)$	$P_M^2(h)$	$F(h)$	$G(h)$
1	1000	4000	1.7×10^{-3}	7.5×10^{-2}
2	1000	2000	0.8×10^{-3}	3.65×10^{-2}

Table 5.2 – Heating parameters

Type	$P_m^3(h) = P_M^3(h)$	$D(h)$	$t_s(h)$	$t_d(h)$
1	600	8	1	100
2	400	6	1	100

Table 5.3 – Washing machine parameters

Class	Lighting type	Heating type	Washing machine type	Maximum power $L(h)$
1	1	1	-	5000
2	2	2	-	2500
3	1	1	1	5600
4	2	2	2	2900

Table 5.4 – Houses parameters

Vital washing machine utility is 0/1: it reaches its maximum value when the appliance is operated. Comfort utility is maximized when the washing machine is operated at its earliest start time (see Figure 3.5). Maximum instantaneous utility values for washing machines are normalized so that all appliances have the same possible maximal utility both vital and comfort (see priority discussion in Section 4.3). This choice is made to have a neutral case where all appliances get the same scheduling opportunities.

Two types of houses are considered for this use case: Classes 3 and 4 (see Table 5.4). Class 4 homes have a better energetic performance than class 3 ones (light and heating characteristics from Class 2 plus less power and time required for completing the washing machine operation). For the performance analysis, we will consider a homogeneous scenario with 100 houses of class 3, and a heterogeneous scenario with 50 houses of class 1 and 50 houses of class 4.

The main characteristics of the appliances are summarized in Tables 5.1 to 5.3 and the composition of each house class is summarized in Table 5.4.

4 Performance analysis

We now present numerical results for each of the supposed use cases. For the rolling blackout scheme, we use a "on" duration of $t_{RB} = 8$ time slots. For the Sub-Greedient problem, we fix the maximum number of iterations to $K_{MAX} = 100$. Two variants are considered: $SG-1$ uses a diminishing step

($a_1 = 1200000$) and $SG-2$ uses a constant step length ($a_2 = 6000$). Parameters a_1 and a_2 were manually tuned (Section 2.2.4).

4.1 Use case: Lights and heaters

We discuss now the performance of the proposed schemes on homogeneous and heterogeneous scenarios.

4.1.1 Results on the homogeneous scenario

Figure 5.3 illustrates the main results for the homogeneous case. It displays the relative utility[6] over the DR period as a function of the available capacity C, for the five considered schemes (GM, LM, RB, $SG-1$ and $SG-2$). For better readability, vital and comfort utilities are displayed separately. We remind that when comparing performance, the lexicographic order need to be taken into account. Indeed, we start by comparing against relative vital utility. When schemes provide the same vital utility, performance is determined by comfort utility value.

Note that a value of interest for vital utility is 0.875, which corresponds to situations where all houses are able to achieve vital light ($P_m^1 = 50$ W) but none has the power necessary for heating ($P_m^2 = 1000$ W) so there is no control of temperature. It can be seen as a floor performance for benchmark.

GM, the optimal solution, achieves maximal vital utility even for very low capacities (down to 3×10^4), thanks to its ability of finding a working rolling allocation that allows all houses to use heat for a sufficient part of the period. Based on GM results, we can measure the gap between optimal allocation and allocations obtained with the partially distributed solutions.

We can see that the performance of RB scheme is terrible for all considered capacities since relative vital utility is below the vital utility threshold 0.875. This is not surprising as when powered off, a house cannot even activate a single light bulb.

Using a static allocation, LM struggles for rising the vital utility above the 0.875 threshold. It can only start to use heat for $C = 10^5$ (1000 W per house). Maximal vital utility is reached for $C = 105 \times 10^3$ (1050 W per house) and maximal utility (vital and comfort) necessarily requires $C = 2 \times 10^5$ (2000 W per house). However, LM achieves descent performance results for high enough available capacity values.

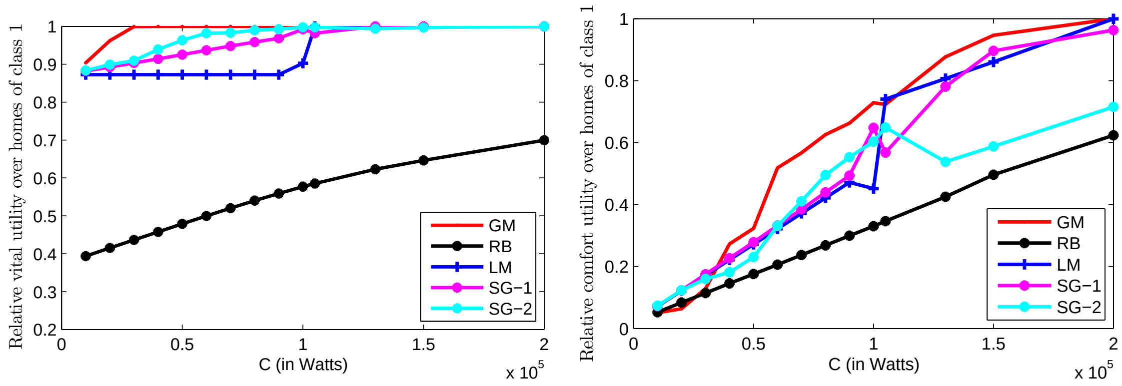
Given the poor performance of RB, we will use LM (along with GM) to analyze the added value of feedback provided by $SG-1$ and $SG-2$.

As expected, our proposal, $SG-1$ and $SG-2$, stands in-between the two opposite schemes GM and LM. It is able to improve the vital utility of houses for values below $C = 10^5$, even if it fails to perform as good as GM. With respect to the comfort utility, it performs on par with LM even in situation where it devotes resources on heating (for vital utility) while LM does not.

It should be noted that the homogeneous case is a kind of worst case scenario for SG. Actually, by design, if all homes have the same $\alpha_k g_{ht}$ for a certain t, SG struggles to break ties between homes. In addition, we can see that using the constant step length method ($SG-1$) seems to be better suited for low capacity values. However, for capacity values for which vital utility are fully provided, the first step size method ($SG-2$) achieves a better performance.

Due to the homogeneous setting, the algorithm consumes many iterations to reach its best solution. It reaches the best found allocation after up to 10 iterations for the first step size method ($SG-1$) in our experiments, while the whole 100 iterations may be needed to reach the last best solution found when considering the second step size. The second method $SG-2$ requires 37 iterations on average to reach the best found solution for the supposed capacity values.

[6]The maximal value 1 is reached when all appliances from all homes of a given class reach their maximal utility.



(a) Vital utility (b) Comfort utility

Figure 5.3 – Relative utility as a function of the available capacity for partially distributed schemes (homogeneous case, class 1)

As we are about to see, SG performs better in a heterogeneous case.

4.1.2 Results on the heterogeneous scenario

The results for the heterogeneous case are shown in Figure 5.4 (class 1 and class 2).

We omit to display *RB* as the heterogeneous scenario adds no insight for that scheme.

We briefly remind discussion on the optimal solution given by GM from Section 7.3. For vital utility, the results are pretty much similar for both classes to the homogeneous case, with maximal value obtained even for low capacities (down to 3×10^4). For the comfort utility, however, one notices that GM leads to better values for class 2 compared to class 1. This is due to the fact that class 2 houses have better energetic performance, so once vital utility is ensured for all, it is more efficient to allocate energy to homes of class 2.

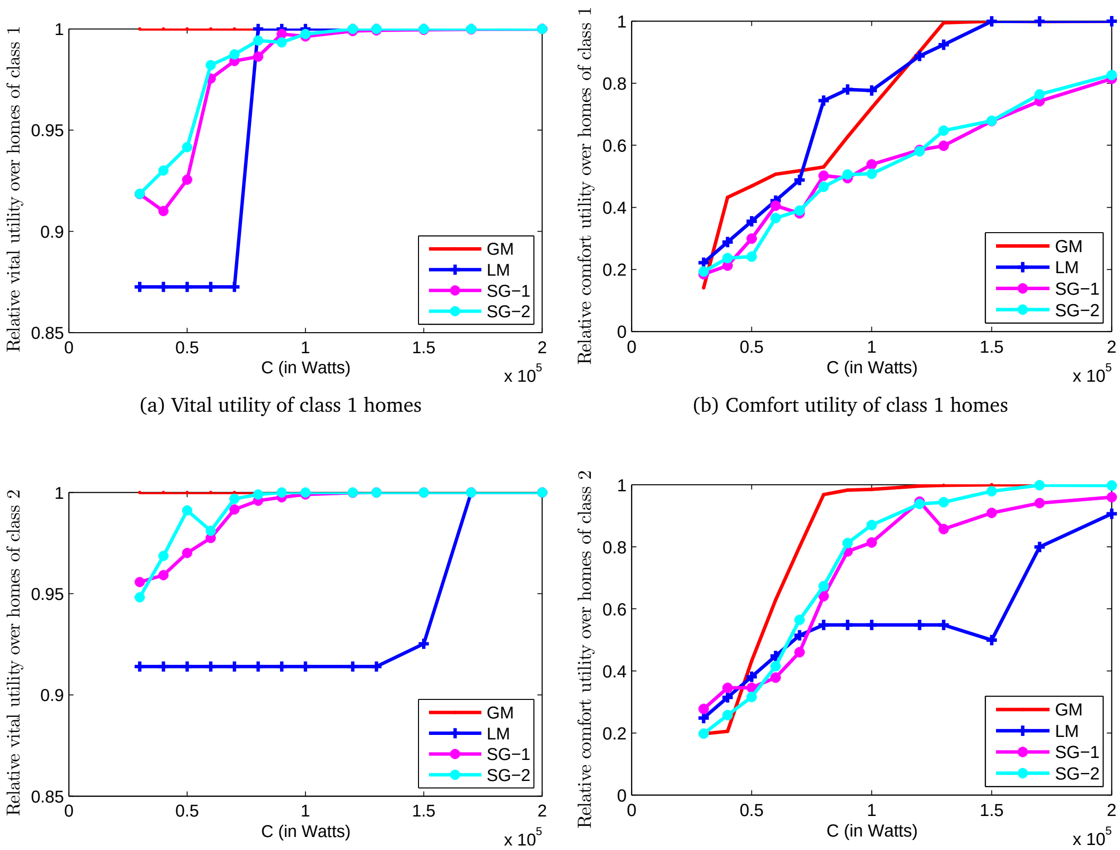
The same reason explains the poor performance of LM. Let us remember that the static allocation is proportional to the maximum power $L(h)$ of homes. So for a given capacity, class 1 homes get more power than class 2 ones. As a result, while performance of class 1 is satisfactory, performance of class 2 is terrible despite the better energy performance of class 2 homes. In particular, the capacity required for class 2 houses to achieve maximal vital utility is very high: $C = 1.7 \times 10^5$, which corresponds to 1700 W per house (regardless the class).

We observe that compared to the homogeneous case, the performance of our solution SG is now closer to GM than to LM. In particular, SG manages to take advantage of the heterogeneity to reach high vital utility values more quickly than in Figure 5.3. Regarding comfort utility, it stays below GM values but manages to give descent values for both classes, which gives a clear advantage over LM (especially regarding the handling of class 2 houses).

Moreover, on the heterogeneous case, SG takes on average 12 iterations to find the best allocation using the first step size method ($SG-1$) and 23 iterations using the second step size method ($SG-2$).

4.2 Use case: Lights, heaters and washing machines

We discuss now the performance of the proposed schemes when program-based loads (washing machines) are added to the mix of appliances.



(a) Vital utility of class 1 homes

(b) Comfort utility of class 1 homes

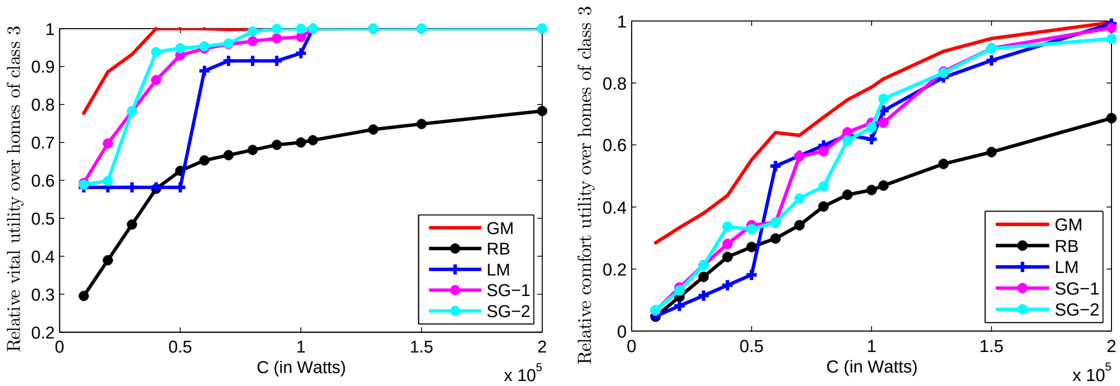
(c) Vital utility of class 2 homes

(d) Comfort utility of class 2 homes

Figure 5.4 – Relative utility as a function of the available capacity for partially distributed schemes (heterogeneous case, classes 1 & 2)

4.2.1 Results on the homogeneous scenario

The main results on the homogeneous case are presented in Figure 5.5. It also displays the relative utility per home over the DR period as a function of the available capacity C, for the five supposed schemes: GM, LM, RB, $SG-1$ and $SG-2$.



(a) Vital utility

(b) Comfort utility

Figure 5.5 – Relative utility as a function of the available capacity for partially distributed schemes (homogeneous case, class 3)

The maximal feasible utility (vital and comfort) is also normalized to 1 for this use case. Another value of interest for vital utility is 0.58, which corresponds to situations where all houses are able to achieve vital light ($P_m^1 = 50$ W) but none has the power necessary for heating ($P_m^2 = 1000$ W) so there is no control of temperature, nor washing machines are scheduled. When washing machine are scheduled in addition to lights (without heating), vital utility reaches 0.92.

We notice that the performance of *RB* is improved compared to the previous case. Indeed, this is mainly due to the utility gained from scheduling washing machines. In this numerical analysis, the rolling blackout duration ($t_{RB} = 8$) provides the required duration of operation of a washing machine. This shows a (marginal) interest of the *RB* scheme when program-based loads are involved.

We see that LM has similar performance as in the previous use case for capacity values that are lower than required power for scheduling all washing machines ($P_m^3 = 600$ W per home). For $C = 0.6 \times 10^5$ till $C = 1.05 \times 10^5$, vital utility is increased compared to the previous use case since washing machines are scheduled. Similarly to the previous use case, LM requires 1050 W for each home to reach the maximum vital utility and 2000 W to reach maximum comfort.

We can expect that adding washing machines will make it more challenging for SG to improve utility in addition to having homogeneity. Indeed, for very low capacity values ($C < 2 \times 10^4$ W), $SG-1$ and $SG-2$ perform slightly better than LM. Then, for capacity values up to $C = 6 \times 10^4$ W, $SG-1$ and $SG-2$ significantly over-perform LM. In particular, the schemes manage to activate must washing machines starting from $C = 4 \times 10^4$ W (400W per home on average, to compare with the 600W required to operate a washing machine). For capacity $C \geq 7 \times 10^4$ W, vital results are actually relatively better than those of the previous considered use case. Similarly, for comfort utility, SG and LM have very similar performance for high capacity values.

As for the number of iterations required to reach the best solution, $SG-1$ takes around 12 iterations on average and $SG-2$ takes 21 iterations. The slowest convergence is observed for capacity $C = 2 \times 10^5$ W where $SG-1$ takes 95 iterations and $SG-2$ takes 98 iterations.

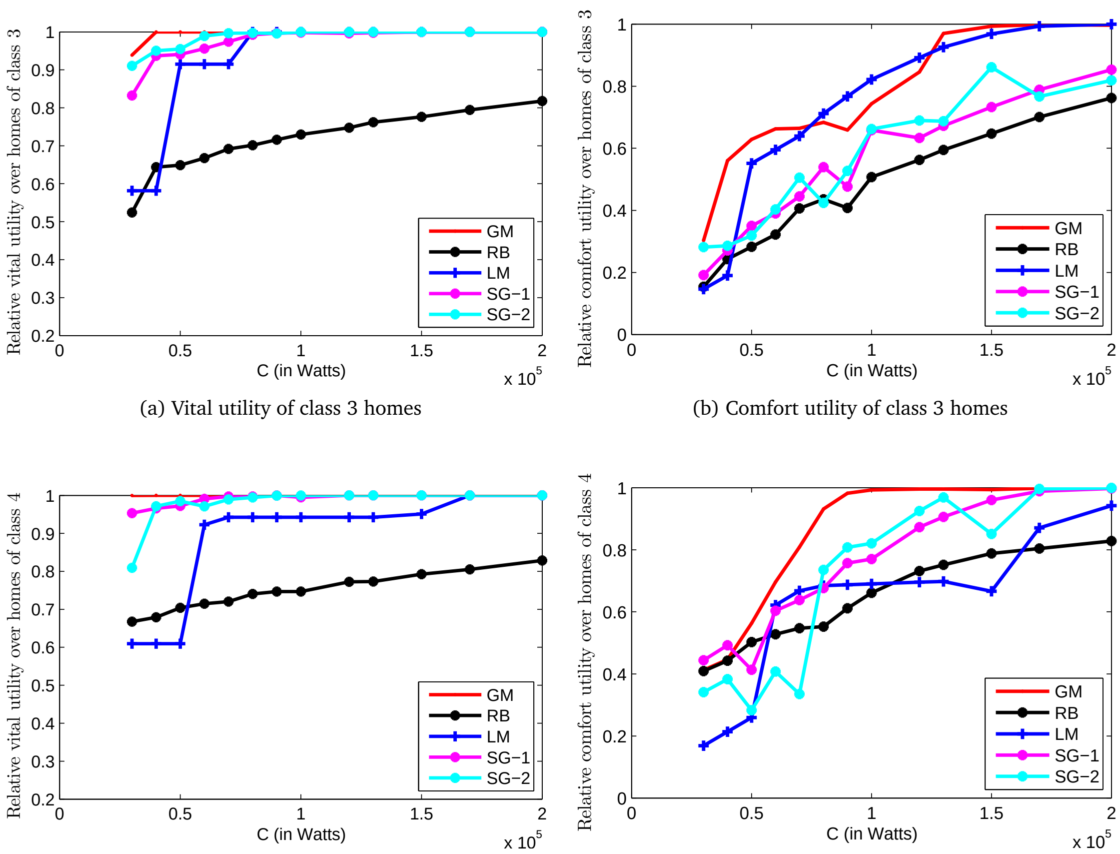
4.2.2 Results on the heterogeneous scenario

Figure 5.6 illustrates the main results for the heterogeneous case for homes of classes 3 and 4. As in the previous case, the effect of adding a washing machine mainly shows for low capacity values.

The performance of RB is improved compared to LM for capacity values for which only RB can schedule washing machines. For LM, since homes capacity is decided proportionally to their static limit, a global capacity $C = 5 \times 10^4$ will only allow homes of class 3 to schedule their washing machines. Actually, for this capacity value, homes of class 3 will get a power limit of 659 W whereas homes of class 4 will only get 341 W (insufficient for turning on a washing machine). For capacity values above $C = 7 \times 10^4$ (corresponding to slightly more than 450 W for each home of class 4), performance is similar to the previous use case since it corresponds to capacity values for which the washing machine get scheduled and minimum lighting requirements are fulfilled.

As for $SG-1$ and $SG-2$, we can see a major improvement in performance compared to LM. Indeed, SG is capable of providing near maximal vital utility for $C = 0.6 \times 10^5$ W whereas for LM maximal vital utility is reached for a capacity $C = 1.7 \times 10^5$.

As for the number of iterations corresponding to the last solution improvement, $SG-1$ takes up to 3 iterations and $SG-2$ takes 14 iterations on average.



(a) Vital utility of class 3 homes

(b) Comfort utility of class 3 homes

(c) Vital utility of class 4 homes

(d) Comfort utility of class 4 homes

Figure 5.6 – Relative utility as a function of the available capacity for partially distributed schemes (heterogeneous case, classes 3 & 4)

5 Conclusion

In this chapter, we proposed a new two-way hierarchical scheme capable of mitigating privacy problems that arise from a centralized solution. Indeed, only coarse-grained dynamic information is made available to the central aggregator which keeps information disclosure to a minimum. It also allows to reduce complexity and thus reach scalability goals by supposing intelligent decision entities at home capable of producing local appliance-related scheduling decisions.

To study the value of two-way communication, we proposed two simple schemes Rolling-Blackout and Local Maximum that rely only on one-way communication. These proposed schemes can be envisioned as potential solutions to deal with capacity allocation problem provided static information from households. In this case, fairness is introduced by balancing the amount of energy provided to homes depending on system capacity. In addition, since the aggregator decision is based on coarse-grained static data, no private user information is disclosed.

Through numerical analysis, we show the high value of feedback in cases where supposed schemes based one-way communication fail to reach the optimum. These cases usually correspond to situations where capacity is very scarce making the cost of inefficiency high in terms of users reduced utility. This shows that our proposed two-way communication provides a good trade-off between improved efficiency due to enabling feedback from homes to express power needs without fully disclosing fine-grained information, and desired scalability feature.

While providing privacy, a two-way communication scheme still requires a central coordinator. In the next chapter, we study schemes in which communication is introduced between homes to enhance the efficiency of power allocations. Indeed, the schemes that we will propose require no or minimal coordination to work.

Chapter 6

Distributed control

In the previous chapter, we showed the benefits of hierarchical scheme with feedback: a trade-off between the performance of a global solution and the privacy offered by one-way schemes. In this chapter, we aim at proposing solutions with better performance on all desired features: scalability, efficiency and privacy. This is achieved by allowing communication between homes. Indeed, provided with anonymity, homes can coordinate with each other to enhance their quality of experience. This eliminates the need for feedback which enforces privacy. This also yields a scheme that is resilient to single point of failure.

This chapter is organized as the following. Section 1 presents motivation and challenges related to the development of a distributed solution. Section 2 presents the related work. System general architecture and specific proposed scenarios are introduced in Section 3. Section 4 presents the distributed solution along with different settings that can be selected. Numerical analysis of proposed schemes is presented in Section 5. Section 6 concludes.

1 Motivation

Like exposed in Chapter 2, DR provides crucial resources for the power system that can help enhance its reliability and support the deployment of renewable energy sources. The abundance of these resources do not only depend on the effectiveness of the control scheme but also on the degree of adoption among energy consumers. So, the popularity of a DR scheme will depend on its capacity to provide some privacy assurance in addition to flexibility and service availability. These requirements are the most significant especially for residential consumers. In addition, communication and computation requirements may be a burden when a large number of homes are involved in the DR solution. Indeed, when a centralized or a hierarchical system with two-way communication is supposed, the need to send dynamic information may require a significant amount data and control signals to be exchanged between decision entities of different hierarchical levels. Furthermore, the need for a central entity to coordinate entities on lower levels of the hierarchy may lead to security threats related with single point of failure which makes the system vulnerable to denial of service attacks.

Security and scalability requirements can be addressed by allowing homes to communicate in order to enhance their quality of experience in a distributed fashion while complying with system constraints. However, different levels of cooperation between homes can be envisioned.

2 Related work

Distributed DR schemes have been explored as a way to mitigate the problem of single point of failure when supposing a central management unit. They insure greater robustness and scalability by limiting information exchange with aggregation levels. They also provide privacy guaranties. Communication between consumers was investigated in the context of pricing-based DR and EV charging. In [69], a scheduling game is proposed among consumers under peak pricing scheme: utility company sends prices to users that depend on the total aggregate load. Prices are decided so that peak to average consumption ratio is minimized. Based on the prices, users cooperate in order to minimize their bills. To do so, each user decides her schedule based on other users aggregate consumption. Scheduling decisions are then broadcast to all playing users. Similarly, a non-cooperative scheduling game is defined in [3] to lower the peak with users owning local generation sources and energy storage devices.

Distributed control schemes are proposed in [74] to study the value of coordination between distributed energy sources inside a home. Coordination is supposed in view of cost minimization objective. They show that depending of the pricing scheme, coordination may or may not be efficient. However, they do not discuss possible coordination techniques and their effect on quality of experience of residents.

In [85], a two-step pricing based DR mechanism is proposed which is very close to our approach. In the first step, a hierarchical system is supposed in which the utility company announces prices that are higher when consumption further deviates from its average value. Based on the prices, users schedule their appliances and save their minimum achievable cost value. In the second step, users can coordinate to flatten their power consumption. Coordination is validated if two coordinating users can both have a similar or a lower cost value. Authors show that by being truthful, users can achieve their best cost. The proposed coordination scheme requires to propagate aggregate load among users which prevents parallelized coordination. In our proposal, we present an approach that allows users to coordinate with any other available user at any time. In addition, we suppose strict constraints on power that can be consumed by users.

Cooperation mechanisms and welfare consensus algorithms are often proposed to deal with EV charging. In [82], cooperation is considered to cap charging consumption of all EVs to a maximum desired limit. The authors propose a real-time scheme where control is applied on time slot consecutively. Given the nature of controlled devices (i.e., batteries), no temporal dependence is supposed between slots, which makes the slot per slot optimization approach justified.

Unlike previously presented efforts that suppose mainly a pricing-based approach, we focus on distributed algorithms for Direct Load Control. Since the algorithm should deal with heterogeneous devices of different residential consumers, a study of possible level of cooperation is needed. In addition, interaction between homes is studied in view of the need for limited communication.

3 System architecture

We consider the general architecture presented in Section 1.1 and we focus on schemes that rely on horizontal communication between decision entities. In the two-level hierarchical architecture, this translates in Demand Response schemes where decision entities at homes interact to enhance the quality of their allocations. DE_h still denotes a decision entity at a home h.

Indeed, we suppose that initial allocations of homes are decided by the aggregator to limit aggregate consumption to a certain desired capacity. These initial allocations describe the maximum power homes may consume. They are discussed to realize a simple environment for the analysis. However, as we will see next, homes will only require the knowledge of their consumption limitations to preserve

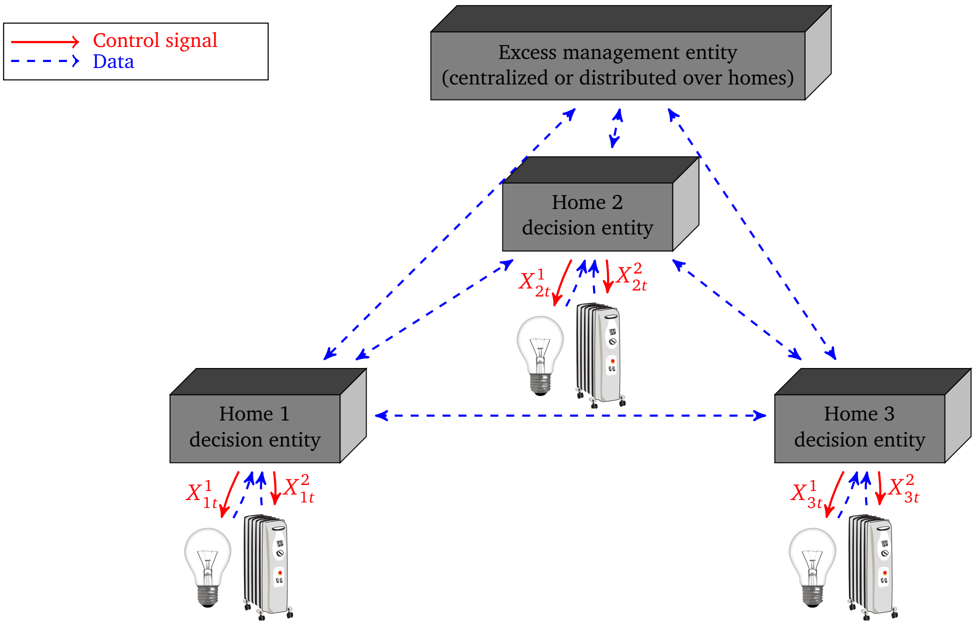


Figure 6.1 – Distributed control

the overall desired capacity. The scheme used by the aggregator to decide these initial allocations can follow any ahead-of-time direct load control mechanism like the ones described in Chapter 5. We suppose that home consumption limitations are imposed for a known finite time period and homes are notified about their consumptions limits before the beginning of the targeted period.

In order to improve the efficiency of their allocations, we consider a “private” marketplace or platform that is provided to homes. The goal of this platform is to enable homes to set bilateral contracts to exchange capacity blocks. Hence, homes contracting the same aggregator can give away capacity blocks (e.g., in case capacity will not be fully used) or get more capacity (e.g., when the home desires to consume more than imposed capacity). Capacity blocks can be of any size (i.e., we do not impose a standard unit size for blocks).

The interaction between decision entities is illustrated in Figure 6.1. This figure shows that home decision entities can exchange information. However, each home can only control its own appliances. So, each of the home decision entities schedules its appliances based on:

- hard limit capacity constraints imposed from the grid side (e.g., by a utility company or an aggregator),
- utility function defined for each appliance,
- information provided by other home decision entities.

Utility functions for appliances follow the model proposed in Section 3.1. Indeed, two levels of utility per appliance will be considered, namely vital and comfort. We remind that the first one expresses high priority targets of high impact on users' wellbeing while the second one expresses less essential preferences.

A simple way to build the exchange platform would be to use a Distributed Hash Table (DHT) architecture [61, 63]. Communication between homes using a DHT platform works as follows. When joining the platform, each DE_h will be attributed an Identifier (ID) from an identifier space. This ID is used to hide the real identity of a home that allows to locate it on the physical power grid. Each home can store a local routing table of a number of successors (e.g., log(H) where H is the total number of homes). Based on the routing table, homes can look up other homes using their IDs. Anonymity can be further enforced by deploying dedicated distributed platforms like Freenet[1] or Tor[2]. Such platform makes it more difficult for data exchanged to be traced back to originating home. Thus, through this platform, homes can interact and exchange information without disclosing their identity. Provided anonymity helps mitigate privacy problems when homes need to exchange information.

In addition to homes decision entities interaction, we also suppose a new functional group that we call excess management entity (see Figure 6.1). This entity is introduced to better deal with unused capacity by homes. Its function may be distributed among home decision entities, it may also be introduced as a separate central component. These two supposed cases are elaborated in Section 4.3.

4 Problem formulation

We aim at proposing schemes that allow homes to exchange power in order to improve the efficiency of their allocations. The basic building block of our scheme is simple to describe: a set of homes tries to enhance their quality of experience by solving a local optimization problem. This central operation is called an *activation*. If successful, an activation will result in a multilateral exchange of capacity blocks corresponding to the solution of the problem. In details, we suppose that when homes use the exchange platform, they can contact either one or more peers in order to explore power exchange possibilities.

Suppose N is the number of homes that will be involved in the coordination. One (or more) of the homes participating to the coordination will solve a social welfare maximization while taking into account information provided by the other $N-1$ homes. Allocation will be determined such that homes will seek to comply with the sum of their capacity limits. The optimization is formulated by equations (6.1).

$$\max_{X^a_{h_i t}, x^a_{h_i t}} \sum_{t=1}^{t_M} \sum_{i=1}^{N} \sum_{a=1}^{A} U^a_{h_i t} \tag{6.1a}$$

s.t.

$$\sum_{i=1}^{N} \sum_{a=1}^{A} X^a_{h_i t} \leq \sum_{i=1}^{N} C(h_i, t) + C_r(t), \ \ \forall\, t \tag{6.1b}$$

$$P^a_m(h_i) x^a_{h_i t} \leq X^a_{h_i t} \leq P^a_M(h_i) x^a_{h_i t}, \ \ \forall\, t, \ \forall\, i, \ \forall\, a \tag{6.1c}$$

$$x^a_{h_i t} \in \{0, 1\}, \ \ \forall\, t, \ \forall\, i, \ \forall\, a. \tag{6.1d}$$

In this formulation, t_M denotes the optimization time period. Equation (6.1b) imposes that coordinating homes schedule their appliances while making sure that their joint consumption do not exceed the available capacity, which is made of two components:

- parameter $C(h_i, t)$ represents the capacity assigned for home h_i at time t
- parameter $C_r(t)$ represents additional available capacity that homes can use (more details below).

Truthfulness of users can be guarantied by requiring more than just one of the involved homes to solve the optimization and compare the consistency of obtained solutions.

[1] https://freenetproject.org/
[2] https://www.torproject.org/

After (6.1) is solved, the solution is proposed to all users of the set. If all accept, the activation is successful and the solution becomes the new current allocation. Otherwise, activation fails and the allocation of users remains unchanged.

The complete scheme simply consists in executing activations multiple times over multiple sets in an asynchronous fashion until an acceptable solution is reached. That been said, we still need to define a few protocol details to have a complete description of the solution :

Local constraints Are users OK to degrade their own quality of experience in order to increase the global quality of experience?

Scheduling How are the set of peers involved in an activation decided?

Excess Management In a given allocation, it is likely that some spare capacity remains. How to use that (i.e., the $C_r(t)$ term in (6.1b))?

4.1 Local constraints

The previously presented problem do not express any users related local constraints for the approval of the exchange. In practice, a user may not be willing to have her own quality of experience lowered. So, additional constraints to the optimization problem need to be imposed.

We will be interested in local constraints that will express three types of users behavior, namely selfish, cooperative and an intermediate. These behaviors are described as follows:

- *Selfish users* will not accept any exchange if it does not improve their perceived utility;
- *Cooperative users* accept the outcome of the optimization if it improves the perceived utility of the group of involved users;
- *Intermediate users* are willing to cooperate when the outcome only affects their comfort utility but are selfish when it comes to vital needs.

We now formalize the additional constraints corresponding to the three supposed behavior cases. These constraints are to be incorporated into (6.1) to ensure the proposed solution will be accepted. We denote by ΔU_{h_i} the change of total utility for home h_i if the activation is successful.

Additional local constraints are expressed in ΔU_{h_i} and are formulated as follows:

- A selfish user h_i will impose the constraint $\Delta U_{h_i} \geq 0$.
- A cooperative user h_i will impose no constraint but the fact that the proposed solution strictly improve the current one ($\sum_{i=1}^{N} \Delta U_{h_i} \geq 0$).
- An intermediate user h_i will impose the constraint $\Delta U_{v h_i} \geq 0$.

These constraints are added to the problem formulation defined by equation (6.1) depending on the considered type of users behavior. An exchange is approved when utility is globally improved for users. In the present work, we assume that all homes have the same behavior[3]. We thus call Selfish Distributed Maximum (SDM), Cooperative Distributed Maximum (CDM) and Intermediate Distributed Maximum (IDM) the schemes that will result from homes coordination in case of selfish, cooperative and intermediate users respectively.

[3] But the framework allows heterogeneous behaviors for example depending on whether the user has signed for a premium service or not.

4.2 Scheduling

In our proposed solution, each home periodically contacts a few other homes. For evaluation purpose, we approximate this behavior by a round robin scheduler, a *round* corresponding to a sequence where each home has proposed an activation.

The choice of the homes that participate to an activation is crucial. In particular, the set size is important for complexity and security reasons. Indeed, we suppose that, when homes coordinate, they need to share their preferences with their peers through the platform. Provided with anonymity, such exchange of information can be done without disclosing the identity of homes. However, if the number of homes increases compared to the number of homes using the platform, the likelihood of learning about other homes increases. In addition, the complexity of computation increases with the number of home (see Figure 4.2b). Furthermore, the lowest the number the more flexible the platform is since communication burden is restrained.

For this reason, we propose to limit the set size to N=3 homes, i.e. when a home initiates an activation, it contacts 2 other homes. To balance exhaustivity and randomness in the set construction, we propose that one of the two is contacted in a round-robin fashion and the other one at random (among all IDs of the other participating homes). We then consider two simple schedulers depending on whether the proposed activation relies on all three homes or only a subset of them.

These techniques are named Direct Activation (DA) and Best Activation (BA).

4.2.1 Direct Activation

As a first simple scheduler, we suppose that all contacted homes participate to the activation. If all homes are satisfied with the outcome of the joint optimization, required modification in power allocations are directly activated. This scheduler is thus called Direct Activation scheduler and denoted by *DA*.

4.2.2 Best Activation

Instead of directly activating power exchange with all contacted homes, activation will only be performed on a subset. For N=3, this gives the following behavior: the initiator of the activation will compute the result of an activation with the first contacted peer and the result of the activation with the second one, and propose the best option (in term of global utility) among these two. This approach is inspired by the *power of two choices* principle [4] and aims at minimizing the size of the problems to solve. This scheduling technique is called Best Activation since a home chooses the best activation opportunity. It is denoted by *BA* .

4.3 Excess management

We consider three scenarios for dealing with unused power, namely local, distributed and centralized excess management. We call residual capacity, the remaining unused part of the allocated limit to homes. Assuming time is slotted, supposed scenarios correspond to different ways of assessing and managing residual capacity at a time slot. We now elaborate on each scenario and discuss how capacity parameters (effective and residual) are updated. In the following, we denote by $X_{h_i t}$ the total power assigned for appliances at home h_i at time t (i.e., $\sum_{a=1}^{A} X^a_{h_i t}$) after a successful activation is performed. We use the notation $C'(h_i, t)$ to denote the new capacity assigned for home h_i at time slot t after the activation in order to differentiate it from capacity $C(h_i, t)$ used as an input parameter of the activation.

4.3.1 Local excess management

As a first option, each home's DE_h can locally manage its residual capacity. This setting is called Local Excess Management scenario (LEM). This means that provided with a power limit, DE_h will schedule its appliances and always keep record of residual capacity. So, when homes coordinate to jointly optimize their appliances' schedules, decisions are made in view of the total (used and unused) power of coordinating homes. In addition, when an exchange is approved, the home that should give away part of its assigned capacity limit, will only transfer the exact amount of power needed by other homes.

Thus, a home which power allocation is lower than its capacity before the coordination (i.e., $C(h_i,t) \leq X_{h_i t}$) will have an updated capacity $C'(h_i,t)$ equal to $X_{h_i t}$. In this case, the added capacity to home h_i should be taken away from homes for which $C(h_j,t) \geq X_{h_j t}$ where $j \in \{1..N\} \backslash \{i\}$. To do so, we start from the home with the smallest unused capacity that satisfies $C(h_j,t) \geq X_{h_j t}$ and we update it's capacity $C'(h_j,t)$ to $max(X_{h_j t}, C(h_j,t) + C(h_i,t) - X_{h_i t})$ and so on[4]. In this case, no additional available capacity is supposed so the value of $C_r(t)$ is always set to zero for all time slots (the unused capacity is directly embedded into the $C(h,t)$ parameters).

The main drawback of this method is that the overall unused capacity for a given time slot may be scattered among multiple homes, which may forbids the scheduling of a demanding appliance even if there is enough available power. This will lead us to the next scenario.

4.3.2 Distributed excess management

In this case, homes only keep useful capacity values for each time slot and they delegate unused power capacity to distributed excess management entities. This is why this scenario is denoted by Distributed Excess Management (DEM). These excess management entities are chosen among homes participating to the exchange platform. Each management entity is responsible of collecting residual capacity for time slots assigned to it when a power exchange takes place. The assignment of a time slot to a specific home that will take on the role of the excess management entity of the time slot, can be done in a uniform random way through the DHT: some ID is assigned to each time slot, and the home in charge of that ID in the DHT becomes in charge of the time slot. To take advantage of aggregated unused capacity for a certain time slot, homes need to connect and coordinate with the home that represents the excess management entity for that time slot. For this scenario, the aggregation of residual capacity makes it more likely to be useful and hence potentially improves the efficiency of power exchange.

Consequently, the update of the assigned capacity at a time slot t depends on whether the home has the role of the excess management entity responsible for the time slot. So, if DE_{h_i} at home h_i is not the entity responsible for time slot t, capacity $C'(h_i,t)$ is set equal to $X_{h_i t}$. In the case home h_i is the responsible for the residual capacity at time slot t, capacity $C'(h_i,t)$ takes the value $C(h_i,t) + \sum_{j \in \{1..N\} \backslash \{i\}} (C(h_j,t) - X_{h_j t})$. Similarly to the previous case, the value of $C_r(t)$ is always set to zero for all time slots as the residual capacity is still embedded in the $C(h,t)$.

This scenario does not deal efficiently with cases where demands may have a strong temporal dependence. Indeed, for certain appliances, fulfilling power needs at a certain time slot imposes the fulfillment of other power needs at other time slots. This is, for example, the case when a minimum duration of operation is required. Dealing with such cases with a distributed excess management

[4]The reason for starting from the lowest unused capacity is to foster the existence of fewer but larger locally unused capacities, which we believe are more adapted than many smaller ones for improving the solutions.

approach may be inefficient since remaining capacities for targeted time slots may be dispersed over different homes.

4.3.3 Centralized excess management

To better manage unfulfilled power needs that span multiple time slots, we suppose a scenario where residual capacity is delegated to a central excess management entity. We call this scenario Centralized Excess Management (CEM). The role of the central entity is to collect residual capacities each time a power transaction is executed between a number of homes. To take advantage of residual capacity, homes can contact the central excess management entity and add its available residual capacity to their current power limits. In this case, the central entity is reserved until the exchange concludes and no other homes can ask for available residual capacity before it is released.

Suppose parameter $C_r(t)$ represent the capacity available at the excess management entity. After an approved joint optimization, capacities $C'(h_i, t)$ takes the value of $X_{h_i t}$ $\forall i$. The available capacity at the excess management entity $C_r(t)$ for time slot t will be updated according to $C_r(t) + \sum_{i \in 1..N}(C(h_i, t) - X_{h_i t})$.

We note that the central component in this case has an important but non critical role. In other words, if we consider a case where this entity becomes unreachable, homes can still exchange capacity values (like in the “Local management scenario”) until connection is reestablished.

5 Numerical analysis

We suppose the same system assumptions, parameters and use cases presented in Section 3. Indeed, we will focus on the evaluation of control schemes in homogeneous and heterogeneous scenarios. Two use cases are supposed: the first use case considers the control of lights and heaters, and the second one adds washing machines to the set of controlled appliances. In addition, we suppose a constant initial global system capacity over the DR optimization period. In this analysis, however, we suppose a DR period of $t_M = 60$ slots (5 hours) instead of $t_M = 100$ supposed in Section 3. This choice is made to allow a thorough analysis of the different possible options for the distributed scheme without affecting the analysis significance.

For distributed algorithms, homes will attempt activations with other homes in an asynchronous manner (see Section 4). This will be repeated until homes consider their allocation satisfactory. So, in a real setting, distributed schemes will run as background tasks that can react to changes in power limits and queries other homes looking for improvement in a gossip-like manner. For the purpose of analysis, we define:

- an *initial system power allocation*: Since distributed schemes allow homes limits to be modified in a way that improves the overall utility while preserving the global capacity constraint, an initial distribution of the available capacity among participating homes (decided by the aggregator) is needed. Any algorithm can be used to determine homes initial capacity limits such that the sum of these limits is equal to the globally available capacity over the DR period. In this numerical analysis, we choose an initial allocation that follows the LM scheme introduced in Section 2.1.1. This choice is made to show how distributed schemes can modify a simple allocation.

- a *stopping condition* for the optimization problem. While the proposed scheme can run continuously as a background task, the settings considered here are static. So, we need a stopping condition: if the relative solution improvement between two consecutive rounds is less than 10^{-2} or after 100 rounds, the scheme stops. This choice is motivated by the desire to explore

Scheme options	Short name	Full name
Local constraints	CDM	Cooperative Distributed Maximum
	SDM	Selfish Distributed Maximum
	IDM	Intermediate Distributed Maximum
Scheduling	DA	Direct Activation
	BA	Best Activation
Excess management	LEM	Local Excess Management
	DEM	Distributed Excess Management
	CEM	Centralized Excess Management

Table 6.1 – Distributed schemes options reminder

possible improvements even if it becomes low while limiting exploration. Indeed, other stopping conditions could have been envisioned that tries to capture specific events (e.g., fulfilling vital needs, reaching a certain comfort goal).

We suppose the centralized scheme GM (see Section 5.1) as a benchmark scheme. This scheme allows us to assess the gap between solution obtained by distributed algorithms and globally optimal one. We also suppose this scheme to measure computation time differences having in mind the exponential complexity in the number of homes (see Figure 4.2b). We also compare with a two-way communication partially distributed scheme $SG-2$ (see Section 4). This allows us to assess the importance of maintaining fine-grained information while distributing computation as opposed to aggregated information feedback.

We now turn to presenting the results of the distributed algorithm for all previously presented local constraints, schedulers and excess management option, and for each of the considered use cases. We tested all the $3 \times 2 \times 3 = 18$ possible variants of the distributed schemes (see Table 6.1) and isolated the settings that have a noticeable impact on th results. Algorithms are differentiated in graph legends based on the relevant settings: local constraints (i.e., CDM, SDM or IDM), scheduler (i.e., DA or BA).

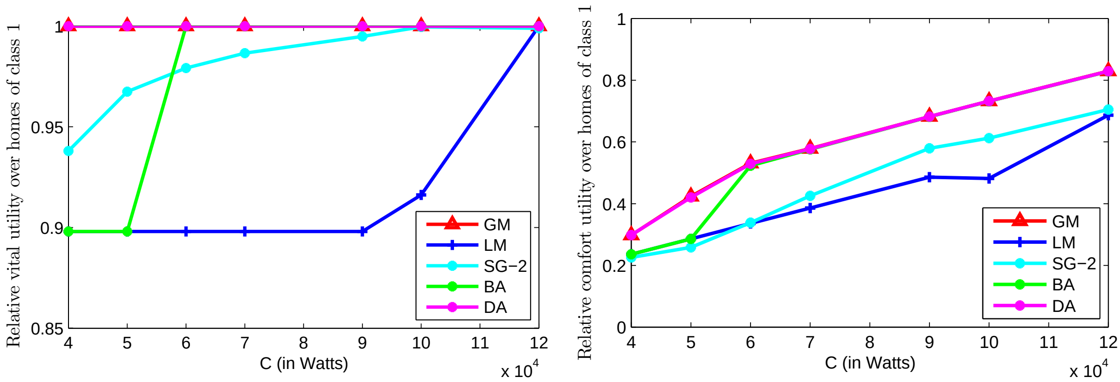
Remark: we do not display on the following figures the effect of the excess management setting because no significant impact appeared in our numerical results. This is a result per se: local or distributed excess management can perform as well as a centralized management, which makes the latter unnecessary. For extremely low capacity values, the effect of a centralized excess management may be more perceivable in cases where program-based loads are controlled and of course residual capacity exists. In the following, we represent results considering a distributed excess management setting.

5.1 Use Case: Lights and heaters

We briefly remind the description of home classes. For the homogeneous case, all 100 homes belong to class 1 and for the heterogeneous one, we suppose 50 homes of class 1 and 50 homes of class 2 (Refer to Table 5.4, page 85 for the specific parameters).

Results of the homogeneous scenario

We analyze the behavior of distributed algorithms in view of different global system capacities. These capacities are chosen so that initial allocation LM cannot provide a globally optimum allocation. As



(a) Vital utility

(b) Comfort utility

Figure 6.2 – Relative utility as a function of the available capacity for distributed schemes (homogeneous case, class 1)

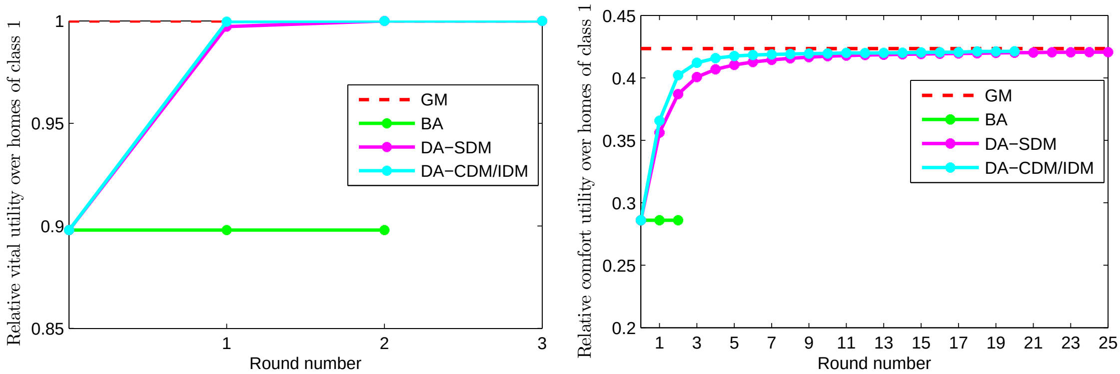
in previous chapters, we plot relative utility that reaches its maximal value of 1 when total maximum utility is reached (either for vital or comfort).

Figure 6.2 presents the relative utility perceived by homes for several supposed global system capacities. We can see that LM and $SG-2$ fail to reach the global maximum for all capacities. As for distributed algorithms, performance reaches that of GM for all tested capacities, the only exception being for schemes using the BA scheduler on low capacities ($C = 5 \times 10^4$W or less). Indeed, for a global capacity of $C = 5 \times 10^4$ which corresponds to 500W per home, optimization over two homes using the BA scheduler cannot guaranty maximum vital. The total power of both homes equals to 1000W and is fully allocated to lighting. So, even if total capacity matches exactly what heaters need to operate, they will not be operated since vital needs for lights will be impacted (lights are seen as more profitable than heaters). In future work, it is interesting to investigate the possibility of borrowing capacity allocated for comfort enhancement in other homes as a new type of “residual” capacity in order to fulfill vital needs.

We propose to look in more details at the case $C = 5 \times 10^4$W. Figure 6.3 provides the evolution of the total vital utility as a function of the number of rounds. We can see that distributed schemes using DA scheduler (which involves 3 homes) are capable of reaching GM performance in less than 20 iterations. However, we can notice that selfishness plays a role in slowing down the convergence of these schemes. We can see schemes using BA stop after the second round since no improvement is perceived (see stopping conditions above).

As for execution time, Figure 6.4 shows the relative running time per round for the tested algorithms. Shown values correspond to a normalization of execution time per round with respect to GM. So, when relative time is equal to 1, the execution time of one round is equal to GM execution time.

We can see that BA is usually faster to solve than DA. This is mostly related to the number of homes involved in the optimization (2 homes versus 3 homes problem to solve). In addition, considering selfish users affects the execution time of a round when schemes use the same scheduler. It may seem that GM has good running time. However, these times are very dependent on the size of the problem as well as on system parameters that affect the complexity of the scheduling algorithm (see Figure 4.2). In addition, since with distributed algorithms computation is similar regardless of the total number of homes, the attractiveness of distributed computation becomes significant when the total number of homes is increased. In addition, in practice, running times for distributed algorithms



(a) Vital utility (b) Comfort utility

Figure 6.3 – Relative utility as a function of the number of rounds for a capacity of 50000 Watts (homogeneous case, class 1)

is actually lower than the time showed on the figure. Indeed, in a realistic setting, during one round, homes can optimize their allocations in parallel.

Furthermore, one can argue that, even if distributed algorithms involve multiple rounds (see Section 4.2), most of the improvement is reached after the first few rounds, so the performance will remain good even if the execution time constraints limits the number of rounds.

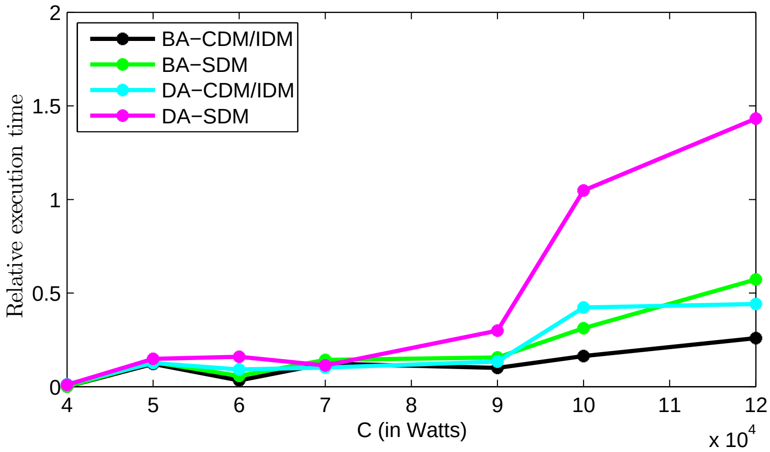
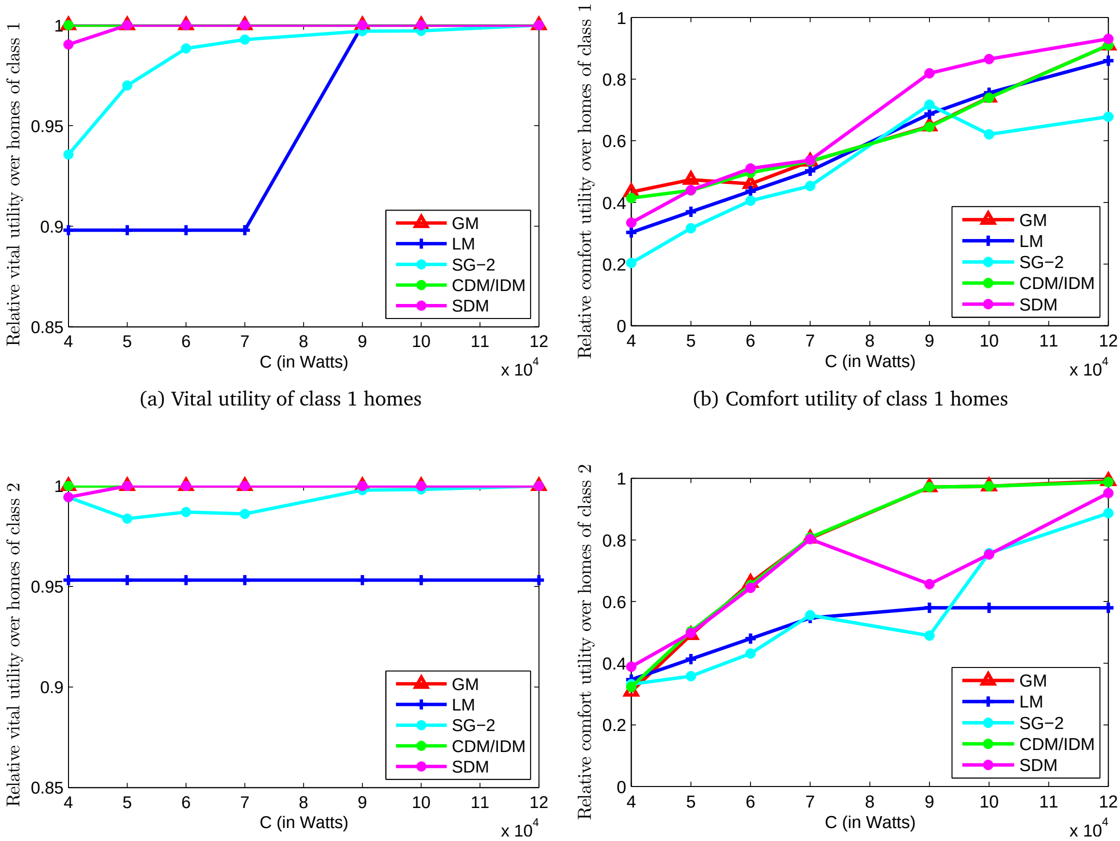


Figure 6.4 – Relative computation time (homogeneous case, class 1)

Results of the heterogeneous scenario

In a heterogeneous scenario, Figure 6.5 shows results for a number of global capacity values. We can observe the same behavior as in the homogeneous scenario where $SG-2$ offers a performance that is better than LM but that may not reach that of GM for all capacity values. In contrast, all distributed schemes that rely on cooperative or intermediate behavior achieve near optimal performance for all tested capacities. Selfish behavior is suboptimal for low capacity values (vital sub-optimality) and



(a) Vital utility of class 1 homes

(b) Comfort utility of class 1 homes

(c) Vital utility of class 2 homes

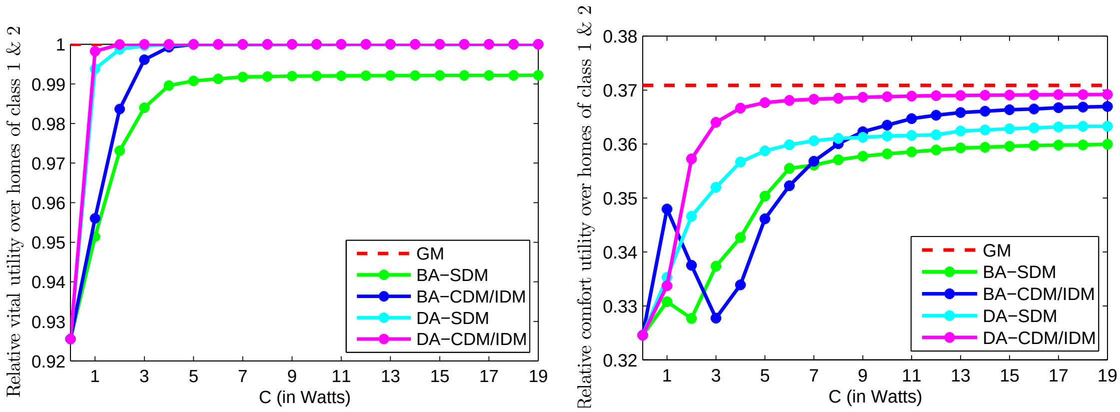
(d) Comfort utility of class 2 homes

Figure 6.5 – Relative utility as a function of the available capacity for distributed schemes (heterogeneous case, classes 1 & 2)

high values (comfort sub-optimality). For example, if we look at high capacities ($C \geq 9 \times 10^4$), we can notice that selfish behavior tend to favor the comfort of class 1 users due to the considered initial allocation. Indeed, having fulfilled their vital needs, homes of class 1 will not sacrifice their comfort in order to enhance that of class 2 homes. This results in a globally suboptimal solution.

Let us now analyze in more details what happens for a scarce capacity.. Figure 6.6 illustrates the convergence speed for a capacity $C = 4 \times 10^4$. We observed 4 distinct types of behavior, depending on the scheduler (BA or DA) and on the selfishness (SDM or CDM/IDM). We can see that heterogeneity enables to escape potential blocking problems depicted for homogeneous homes when scheduler *BA* is supposed (Figure 6.2b). However, we can see that selfish behavior always renders a suboptimal solution compared to not fully selfish one for both schedulers. Actually, a closer look of BA schemes on rounds 2 and 3 shows that non-selfish schemes can more easily trade comfort utility for vital utility. In addition, we can still see that *DA* has a faster solution improvement than *BA*.

Figure 6.7 examines running times of the schemes compared to *GM*. We can see a similar effect than that observed in the homogeneous scenario. Indeed, optimization over 2 homes (*BA* scheduler) has a faster time per round than optimization over 3 homes (*DA* scheduler). In addition, selfishness contributes to a significant increase in the time required to complete a round. A notable phenomenon in Figure 6.7 is the “explosion” of the relative running time for high capacities, more visible than the one from Figure 1.4. We do not have a clear explanation for that behavior and can only suppose an increased relative complexity of activations in the case of high capacities.



(a) Vital utility (b) Comfort utility

Figure 6.6 – Relative utility as a function of the number of rounds for a capacity of 40000 Watt (heterogeneous case, classes 1 & 2)

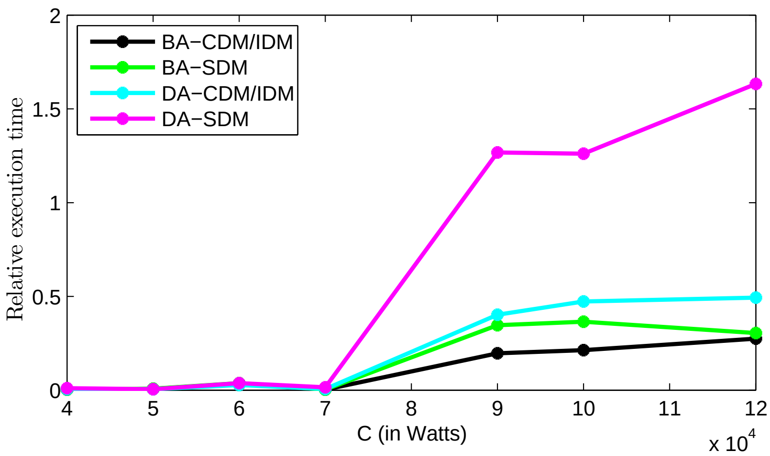


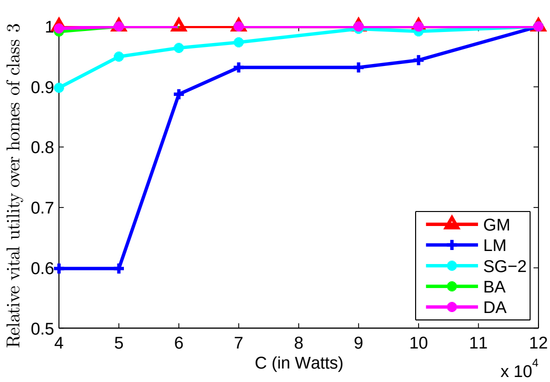
Figure 6.7 – Relative computation time (heterogeneous case, classes 1 & 2)

5.2 Use Case: Lights, heaters and washing machines

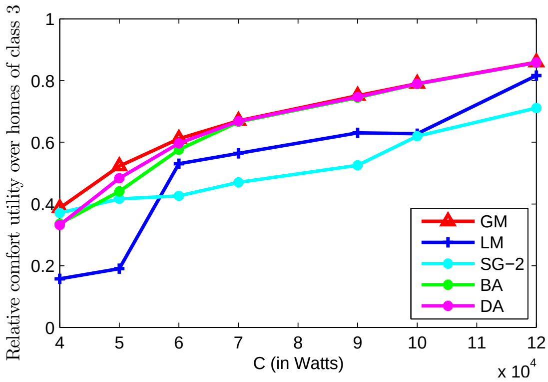
In this use case, we add washing machines to heaters and Lighting which adds time dependence to the allocations since its operation is non interruptible (see Section 4.3). We discuss the two scenarios: homogeneous and heterogeneous. For the homogeneous case, all homes belong to class 3 and for the heterogeneous one, we suppose 50 homes of class 3 and 50 homes of class 4 (see Table 5.4).

Results of the homogeneous scenario

Figure 6.8 illustrates results for several global available capacity values. We can see that for these capacities, the gap between the global optimum given by *GM* and distributed algorithms is very low and only noticeable for scarce capacities. For all supposed capacity values, all distributed algorithms are capable of reaching the global optimum for vital utility.



(a) Vital utility



(b) Comfort utility

Figure 6.8 – Relative utility as a function of available capacity for distributed schemes (homogeneous case, class 3)

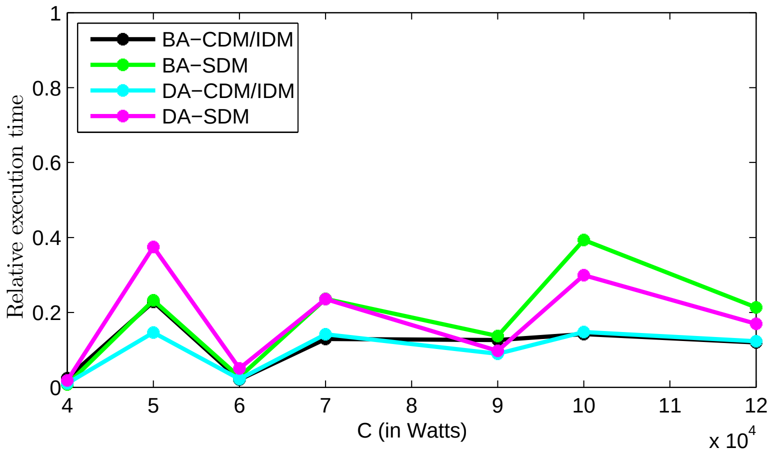


Figure 6.9 – Relative computation time (homogeneous case, class 3)

As for convergence speed, conclusions are very similar to the previous use case. For short, schemes considering a *DA* scheduler have faster convergence compared to schemes using *BA*. Figure 6.9 show the computation of all the schemes for several values of the available capacity. We can see the improvement in relative computation time of distributed algorithms with respect to *GM* when washing machines are added. Indeed, the introduction of washing machines increases the complexity of the problem due to time dependence. While the increase in complexity creates a significant increase in the execution time of *GM*, it has very minor effect on distributed algorithms.

Results of the heterogeneous scenario

The observations for the heterogeneous scenario are very similar to previously investigated cases. Computation time for this scenario is shown in Figure 6.10. Like in Figure 6.9, we can see a significant improvement in computation time of distributed algorithm compared to the previous use case without washing machines. With the increased complexity of the optimization problem, distributing computation reduces significantly the execution time.

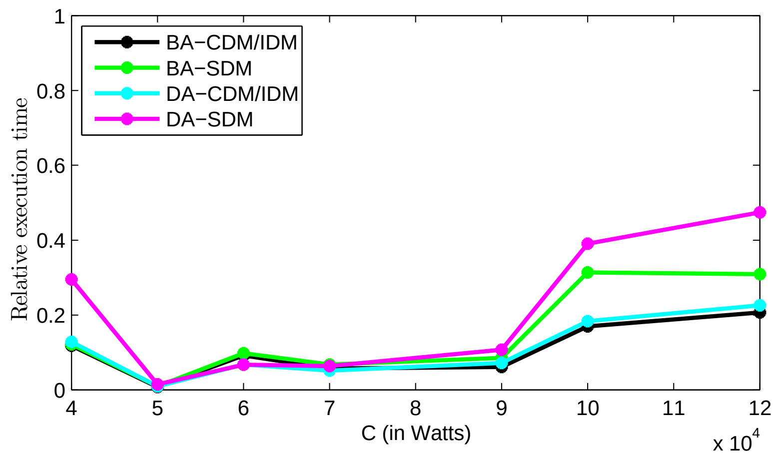


Figure 6.10 – Relative computation time (heterogeneous case, classes 3 & 4)

6 Conclusion

In this chapter, we proposed distributed schemes based on horizontal communication between entities at the same level of the control hierarchy. These schemes allow to alleviate vertical communication burden between entities at different hierarchical levels. Privacy related problems are mitigated by two mechanisms: first, anonymization is enforced by using a DHT-like platform so users are hidden by anonymous IDs. Then, personal information is not disclosed to a central entity, which makes more difficult any attempt of de-anonymization. To analyze the effect of users behavior on the performance of the control, we considered three aspects: selfish , cooperative and intermediate. These behavior aspects are translated into local constraints added to the distributed optimization problem. We analyzed two selection techniques that homes may use to contact other homes on the exchange platform: one based on direct activation supposing 3 homes coordination and one based on the best outcome activation considering two homes coordination. To deal with unused capacity, we analyzed three scenarios of excess management, namely local, distributed and centralized.

Through numerical analysis, we show that distributed schemes allow to reach near-optimal solutions for all tested values of available system capacity while also providing enhanced privacy and better complexity. Indeed, complexity is significantly reduced and highly controlled compared to a centralized solution since smaller instances of fixed size need to be solved. While further deep analysis is required to validate the present observations, we were already able to extract several insight related to the tuning of distributed schemes:

- Unused capacity can be managed in a decentralized way with no significant impact on the result compared to a centralized management.
- An approach with three houses is better than an approach that selects the "best of two choices" in a two-houses optimization.
- While complete selfishness has a negative impact, partial collaboration seems to be enough to get optimal results.

To conclude, proposed distributed schemes are very promising solutions to implement in a real system because of the flexibility it offers to users, the privacy provided, the near-optimal performance and the low complexity.

Chapter 7

Demand dispatch

In the present chapter, we aim at studying direct load control techniques that rely on randomised local control. We propose here a framework distinct from the one used in the other chapters: as opposed to the deterministic approach supposed previously, in this chapter, we suppose that loads have more freedom for taking decisions based on the control signal they receive from the aggregator. These loads are controlled by the aggregator in order to provide Ancillary Services (see Section 2.1) to the grid (i.e., through real-time DR). This is done by requiring the aggregate consumption of loads to follow a power signal sent by the grid operator. The goal of this work is two-fold. First, we assess the impact of poor performance on system stability. This is done by studying the effect of poorly performing resources on other accurate ones (e.g., flexible generators or loads that provide full state information to aggregators). Secondly, we asses system-wide cost in terms of the effectiveness of the tracking under system stability hypothesis.

The chapter is organized as follows: Section 1 present motivation and related work. Section 2 contains the details of the three components that make up the supposed power system control model: the physical grid model, the compensator block and resources actuation block. We focus on the actuator block which includes the models of the classes of loads involved in the DR mechanism. In Section 3, we provide an instantiation of the actuation block where two classes of loads are supposed along with traditional ancillary service resources. Conclusions are presented in Section 4.

This chapter describes joint work with University of Florida and Inria published in [MKBM16]. It exposes results mainly credited to University of Florida. My contributions are mostly related to understanding of power grid control model.

1 Motivation and related work

The existence of resources capable of providing ancillary services to the grid is crucial to maintain system stability and compensate imbalances in generation and consumption. It is demonstrated in [9, 101]) that adopting performance payments can increase the availability of fast responsive resources along with improving quality of control. This implies the added value of high-performance. However, the power grid seems remarkably reliable even though traditional sources of ancillary services do not provide perfect performances. Indeed, Kirby in [49] shows that ancillary services provided by coal generators can sometimes incur significant delay in response. One goal of this research is to understand the costs introduced by poorly performing resources that can explain the resilience of the grid. However, the ultimate goal is to propose control schemes that can provide needed ancillary services from “intelligent loads”.

We focus on DR mechanisms that can provide continuous automatic control. This means that control decisions optimality are tested supposing an infinite horizon (i.e., infinite optimization period). This approach is referred to by *Demand Dispatch* [15]. To engage consumers in participating to this type of services, we envision that incentives can be provided by proposing contracts like the Florida OnCall program. So, a a fixed payment for participation can be provided and regular variable payments that depend on the services actually delivered to the system (e.g., [101]).

To enable the control of heterogeneous loads, we suppose frequency decomposition of the power signal that need to be tracked [38, 59, 64, 8]. This is based on the decomposition of this signal into several signals having different frequency components that can be tracked by each class of appliances. Deciding frequency components that can be addressed by a class of appliances is based on its natural operation cycle duration. This allows to provide high-quality ancillary services without impacting the quality of service delivered to consumers.

To produce control decisions for each load belonging to a given class, we use a model based on the one developed in [64] that targeted a homogeneous collection of loads. The model exploits the large number of resources in order to define randomized control policies at each intelligent loads such that the aggregation of loads are capable to create the global desired power consumption curve.

The highest time-scale of ancillary services that can be provided by proposed control is on time-scales corresponding to primary reserves. However, poor performance on these time-scales can yield the highest risks. Risk is lowered when control time-scale ranges from several minutes to hours. Using arguments from classical control, we study the value of accuracy and performance based on the frequency bands of the control signal. To do so, we propose a power grid control model. We now elaborate on related efforts that target grid and control modeling. Indeed, models that can capture the dynamics of the entire grid have been extensively studied, and have received attention recently due to concerns about grid inertia. The recent work [21] studies the impact of new trends, such as the deployment of inverter interfaced generation on the dynamics of the grid. The paper [76], following [22], shows the diverse effects that fast resources can have on system stability. These grid models are used in our investigations here.

2 A Distributed Control Problem

The analysis is based on a standard input-output model of the grid [52], denoted G_p in Figure 7.1. The “desired behavior” will be zero, and $Y(t)$ will represent a deviation whose desired value is zero. The actuation block in the figure is obtained using demand dispatch along with conventional sources of ancillary service.

The realization of desired power signal can be obtained using a hierarchical control strategy. The ‘intelligence’ at each load involves local control at each device to achieve two goals: reliable service to the grid, and high quality of service to the customer that the device serves. The main goal of this

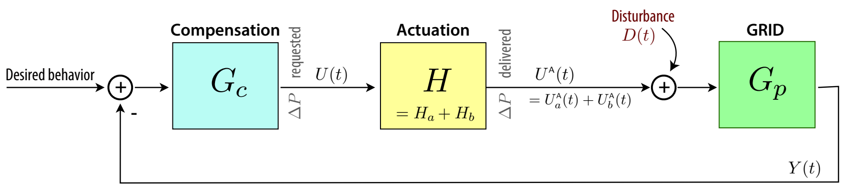


Figure 7.1 – Power Grid Control Loop. We address the question:*where do we find H?*

work is to understand required level of control performance that impact the reliability of the service delivered to the grid.

We begin with a model of the grid.

2.1 Grid dynamics

A transfer function model of the grid is denoted G_p in Figure 7.1. We assume here that the input to G_p is power deviation, and the output is the deviation of grid frequency from its nominal value (60 Hz in the U.S.). In practice, we would also measure tie-line error and other grid disturbances.

An approximate grid model can be obtained by considering an interconnection of aggregate generation and governor dynamics, as in [22] and other papers. The final model is justified by observing the dynamics of the grid following a fault; the initial trajectory can be interpreted as the transient portion of a step response.

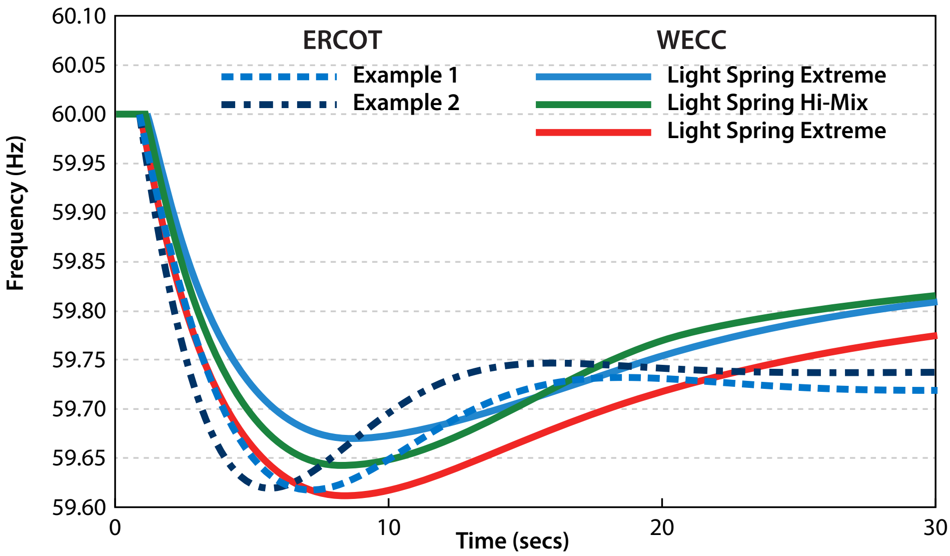


Figure 7.2 – Frequency deviations following a generation outage – examples from WECC and ERCOT.

Figure 7.2 shows examples from WECC (Western US; taken from Fig. 57 of [65]) and ERCOT (Texas) [76]. The slow recovery of grid frequency to its nominal value in the WECC plots is due to secondary control. This is not a part of G_p since we do not consider contingency reserves in this paper.

In theoretical models of the grid, the transfer function depends on the nominal load. This is consistent with observations of the grid following a fault [52, 22]. Consequently, there is significant model uncertainty and variability that must be respected in control design at the grid level.

One version of the ERCOT model is used in the simulation studies summarized in Section 3: the grid model of [22] was used in which the net load (load minus wind generation) is 25 GW, resulting in:

$$G_p(s) = \frac{0.644s + 0.147}{s^2 + 0.4797s + 0.147} \tag{7.1}$$

Corresponding grid frequency dynamics are shown in Example 1 of Figure 7.2. This transfer function has natural frequency $\omega_n = 0.3834$, and damping ratio $\zeta = 0.6256$. The large frequency excursion seen in Figure 7.2 is due to the zero in G_p.

2.2 Actuator dynamics

The transfer function H will be the sum of several transfer functions whose dynamics are shaped by the dynamics of the resource providing ancillary service and some pre-filtering — both locally and at the grid-level. Consider the case of PJM in which H_a and H_b can be associated with RegD and RegA, respectively. Because accurate tracking is required for RegD, the transfer function H_a is shaped entirely at PJM, and the signal $U_a^{\blacktriangle}(t)$ will accurately track a multiple of the RegD signal. The transfer function H_b will include a model of the dynamics of the resources providing the service.

Based on industry practice, and to simplify control design, the following convention is imposed on the feedback architecture: the signal $\boldsymbol{U}$ in Figure 7.1 is scalar valued. Subsequently, it may be decomposed into the sum of several signals differentiated by bandwidth, but these dynamics are modeled in H.

In prior work [64] using two-way communication combined with local control, the loads could track the desired regulation signal nearly perfectly. Without two-way communication, the loads will introduce both dynamics and uncertainty in the transfer function H.

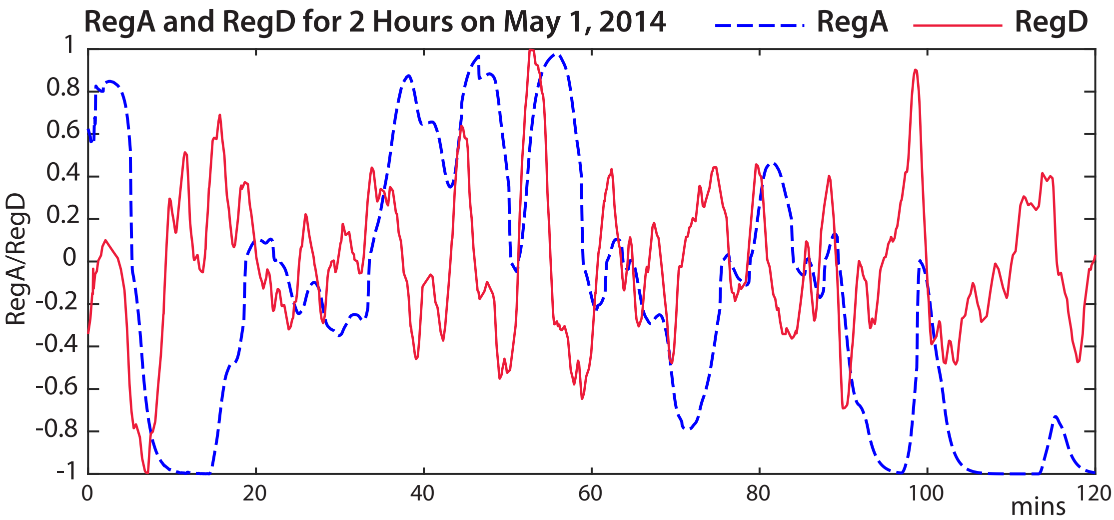


Figure 7.3 – RegA and RegD at PJM for two hours in May, 2014.

We are, however, free to introduce additional local control to improve grid-level performance. Section 3 contains examples in which load dynamics exhibit resonance and some phase-lag. Let G_l denote a transfer function that models the dynamics for a particular class of loads. These dynamics are easily identified, so additional filtering at the load is used to improve the input-output behavior. The dynamics of the filtered loads are expressed

$$H_l(s) = M_l(s)G_l(s) \tag{7.2}$$

The pre-filter M_l is designed so that the collection of loads has reduced resonance, improved phase response, and greater bandwidth.

This design step has not appeared in prior work. The overall design is illustrated through examples in Section 3.

2.3 Grid disturbance

Before turning to the design of the compensator G_c, we first consider how PJM defines the signal $\boldsymbol{U}$ appearing in Figure 7.1. It is obtained by first constructing the ACE, which is a linear combination of grid frequency deviation and tie-line error. The ACE signal is passed through a PI compensator (their G_c) [27], and this is then transformed into a sum of two signals, RegA and RegD. Figure 7.3 compares the two signals over a two hour time period. The higher frequency content in RegD is evident from the figure.

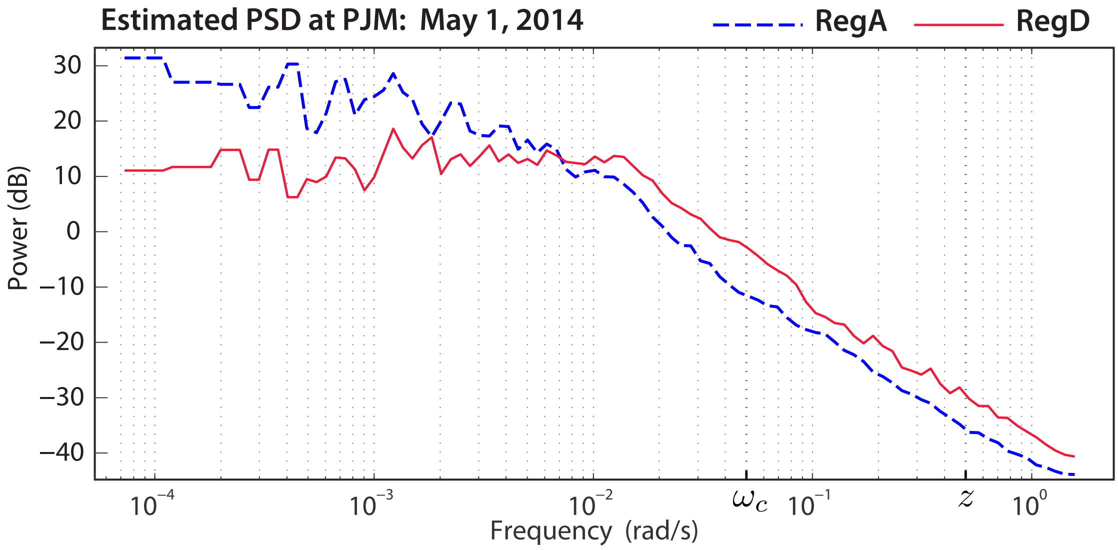


Figure 7.4 – PSD for the RegA and RegD at PJM based on 24 hours of data.

Power Spectral Density (PSD) estimates were obtained using time series over a 24 hour period on the same day. A comparison of the plots shown in Figure 7.4 shows again the higher frequency content in RegD. It also shows very little energy in frequencies greater than 10^{-1} rad/s.

It follows that a compensator G_c at PJM will require high gain for frequencies as high as about 10^{-2}, and relatively low gain at frequencies above $\omega = 10^{-1}$ rad/s.

The balancing reserves at BPA have significantly greater low frequency content. At ISO/RTOs, this low frequency regulation is obtained in the real-time market.

2.4 Control design

The *loop transfer function* associated with the feedback loop in Figure 7.1 is the product of the three transfer functions: $L(s) = G_c(s)H(s)G_p(s)$. The crossover frequency is by definition the value for which $|L(j\omega_c)| = 1$. This frequency is unique by design, and chosen by the grid operator through the choice of parameters in the compensator G_c. The grid transfer function G_p is approximated by a second-order transfer function, whose natural frequency ω_n depends on load. Because of this uncertainty, the compensator should suppress the gain of the loop transfer function in this range of uncertainty (ω_n ranges from 0.2 to 1 rad/s in the ERCOT model).

Following standard design for power systems and elsewhere, we adopt the PI compensator,

$$G_c(s) = K\frac{s+\beta}{s} \tag{7.3}$$

In the simulations described later in the paper, we take $\beta = 0.5$ and fix the gain K so that $\omega_c = 0.05$ rad/s; the crossover frequency ω_c is known to approximate the closed loop bandwidth. With this design we hope to achieve two goals: disturbance rejection in the frequency range where disturbances are present, and robustness to uncertainty in grid dynamics.

2.5 Cost of heterogeneity

Suppose that we have two sources of ancillary service: the first provides perfectly accurate service, but is costly. The second is free, but inaccurate. Our control solutions are insensitive to un-modeled dynamics, especially at low frequency. We may ask, does the introduction of the poor-quality service create a cost, because the more expensive services must work harder?

To address this question we introduce two transfer functions G_a and G_b, modeling accurate and inaccurate resources, and for $\rho \in [0,1]$ denote

$$H = H_a + H_b = (1-\rho)G_a + \rho G_b \tag{7.4}$$

In this subsection we are concerned with the magnitude of service delivered by the accurate actuators, denoted $U_a^{\blacktriangle}(t)$ in Figure 7.1; in transfer function notation, $U_a^{\blacktriangle}/U = (1-\rho)G_a$. The steady-state analysis is based on the following representation:

Lemma 2.1. *The transfer functions from the disturbance to output, and from disturbance to $U_a^{\blacktriangle}$ are given by, respectively,*

$$\frac{Y}{D} = \frac{G_p}{1+L} \tag{7.5}$$

$$\frac{U_a^{\blacktriangle}}{D} = -(1-\rho)\frac{L_a}{1+L} \tag{7.6}$$

in which $L = G_c H G_p$ is the loop transfer function, and $L_a = G_c G_a G_p$.

Proof The first identity is obtained using standard arguments: from the figure we obtain $Y = G_p(D - G_c H Y)$, since "desired behavior" is zero by definition. Solving this algebraic equation gives (7.5).

The following transfer function is easily identified from the figure and the definition of $U_a^{\blacktriangle}(t)$:

$$\frac{U_a^{\blacktriangle}}{Y} = -(1-\rho)G_a G_c \tag{7.7}$$

The transfer function (7.6) is obtained on combining (7.5) and (7.7). □

Consider a steady-state setting in which the disturbance $\boldsymbol{D}$, shown in Figure 7.1, is purely periodic: $D(t) = \sin(\omega_0 t)$ for some $\omega_0 > 0$, and all other signals are periodic with the same frequency. Applying (7.6), we must have $U_a^{\blacktriangle}(t) = k_0 \sin(\omega_0 t + \phi_0)$, where $k_0 = |U_a^{\blacktriangle}/D(j\omega_0)|$ and $\phi_0 = \angle U_a^{\blacktriangle}/D(j\omega_0)$. The gain $k_0 = k_0(\rho, \omega_0)$ is interpreted as the cost; this is motivated by current mileage payments at ISOs [101].

Normalize the transfer functions so that $|G_a(j\omega_0)| = |G_b(j\omega_0)| = 1$. The first resource is assumed to be perfect, $G_a(s) \equiv 1$, and we denote $\phi_B = \angle G_b(j\omega_0)$.

Figure 7.5 shows the cost k_0 as a function of ρ for a range of values of ϕ_B, and a fixed value of ω_0. When there is no heterogeneity ($\phi_B = 0$), the cost decays linearly. For $\phi_B > 90$ degrees, the cost is maximized near $\rho = 1/2$.

The common slope at $\rho = 1$ can be identified as,

$$\frac{d}{d\rho} k_0(\rho, \omega_0)\Big|_{\rho=1} = -\left|\frac{L_a(j\omega_0)}{1 + L(j\omega_0)}\right|$$

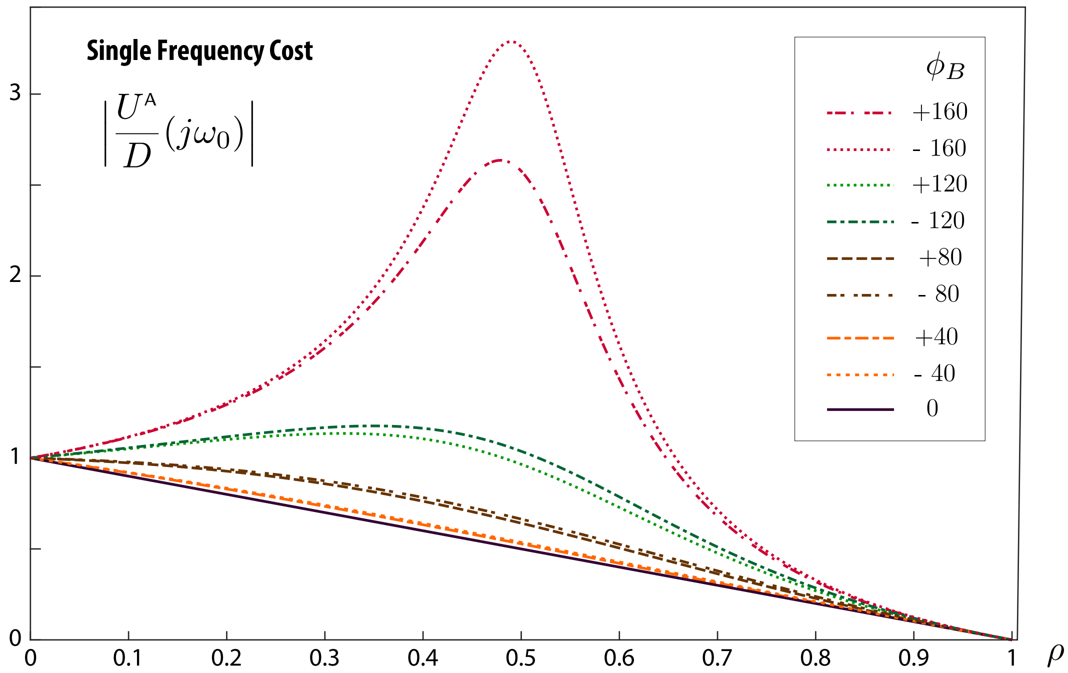
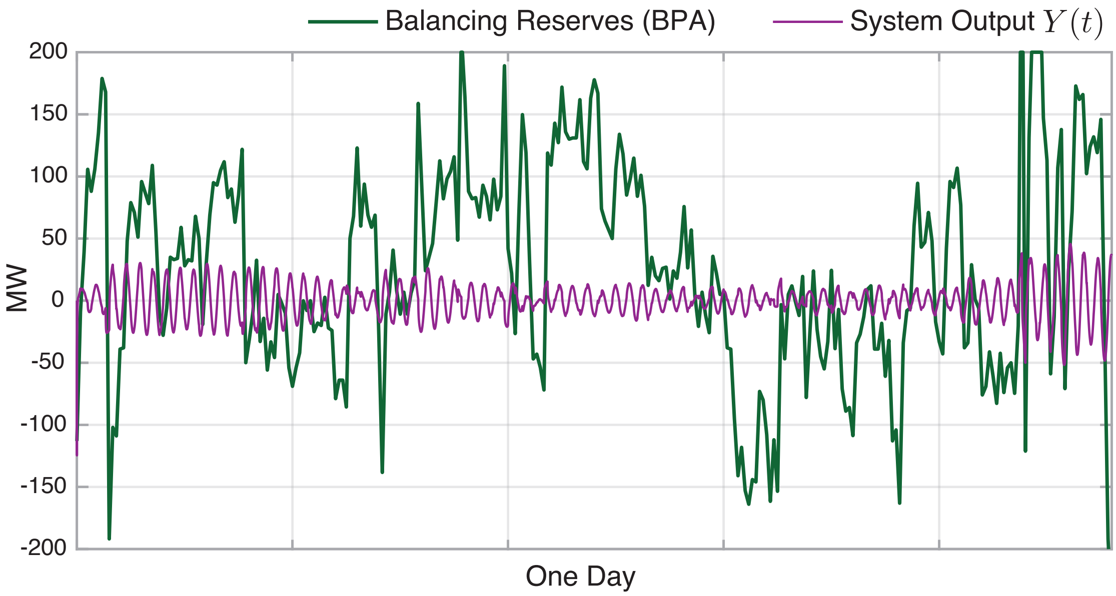


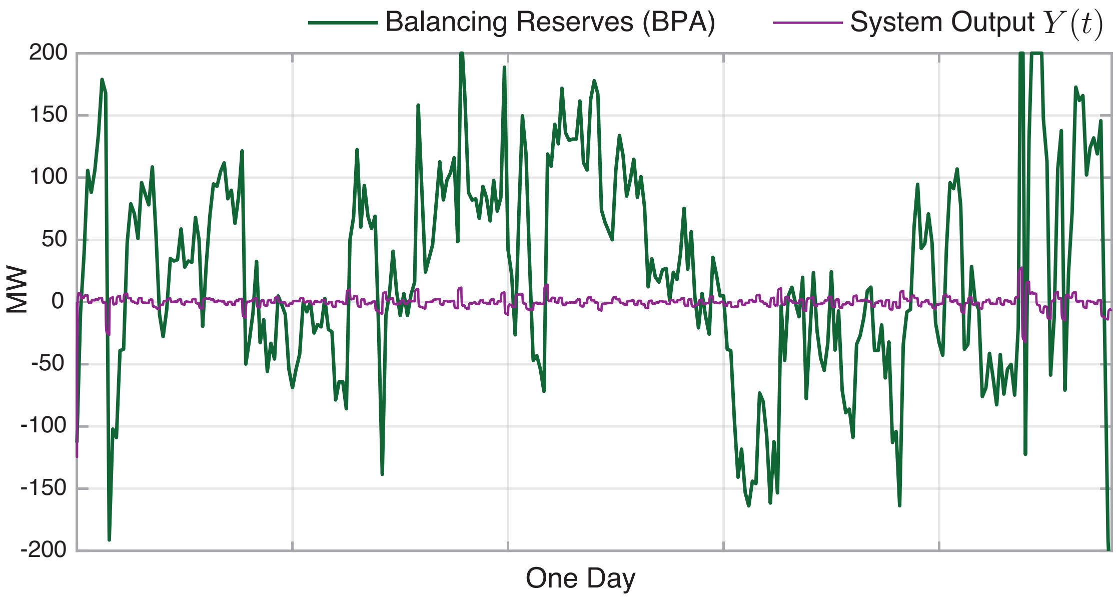
Figure 7.5 – System cost at a single frequency: the highest cost is expected with a mix of approximately 50% of the two sources of ancillary service.

In conclusion: a free resource may be costly if it does not accurately follow the regulation signal. However, the common slope at $\rho = 1$ means that cheap resources are valuable, provided they aren't mixed with others.

All of these conclusions presume that the system is stable. In the next section we consider both stability and cost in a more realistic model.



(a) Lead design with $\omega_{co} = 0.007$ rad/s.



(b) Inverse design, with $\omega_{co} = 0.013$ rad/s..

Figure 7.6 – Comparison of disturbance rejection performance obtained with two different pre-filter designs.

3 Risk & Cost using Demand Dispatch

Experiments were conducted using demand dispatch, with one-way communication from grid to loads.

The grid transfer function G_p in Figure 7.1 was taken to be the second-order linear system (7.1), whose natural frequency is $\omega_n = 0.3834$. As discussed in the Section 2.1, the grid transfer function is sensitive to load and generation mix. This uncertainty means that we should maintain a closed loop bandwidth significantly below the natural frequency of the grid. The PI compensator (7.3) is designed to set the crossover frequency to $\omega_c = 0.05$ rad/s.

The disturbance $\boldsymbol{D}$ was defined as the balancing reserves deployed at BPA during a typical week: June 6, 2015 to June 12, 2015 (data available at [13]). Figure 7.6 shows a plot of $\boldsymbol{D}$ during the first day of this week, along with the closed loop response $\boldsymbol{Y}$ obtained using two different designs. The poor performance seen in Figure 7.6a is eliminated with better local control at the loads. These results are explained in Section 3.5.

The actuator transfer function H is described next.

3.1 Actuator block

The actuator block H in Figure 7.1 is obtained from a combination of three resources: a transfer function G_a representing an expensive, high-quality resource, and two actuators H_{pl} and H_{tcl}, which are based on two separate collection of loads. The two classes of loads work in parallel: the 'pools' could represent a large collection of residential pool loads, or other loads that provide flexibility in a low frequency range. The TCLs represent flexible loads providing ancillary service in a higher frequency range; these may include refrigerators, water heaters, and air-conditioners.

Just as in Section 2.5, the performance of G_a is assumed to be perfect, with $G_a \equiv 1$.

In addition, we introduce a parameter ρ, a low pass filter H_{LP}, a high pass filter H_{HP}, and define the overall actuator transfer function as follows:

$$\begin{aligned} & H = H_{LP}H_\ell + (1 - H_{LP})G_a \\ \textit{where} \quad & H_\ell = H_{pl} + H_{HP}[(1-\rho)G_a + \rho H_{tcl}] \end{aligned} \tag{7.8}$$

This design choice is based on several considerations:

(i) The low pass filter H_{LP} is introduced because of risk of instability due to gain or phase uncertainty from load response: it is designed so that the response from loads is not significant near the crossover frequency ω_c.

(ii) The high-pass filter H_{HP} is used to limit the low-frequency content of the regulation signal sent to the TCLs. This is to help guarantee Quality of Service (QoS) constraints for these loads.

(iii) The purpose of the parameter ρ is to investigate cost as a function of the proportion of TCLs engaged. It will be seen that their imperfect response may introduce cost to the system.

For H_{LP}, we used a second-order Butterworth low pass filter with cut-off frequency ω_{co} and damping ratio $\zeta_{lp} = \sqrt{2}/2$, with transfer function $H_{LP}(s) = \omega_{co}^2/(s^2 + 2\zeta_{lp}\omega_{co}s + \omega_{co}^2)$. The cut-off frequency ω_{co} is a parameter used in our cost/risk evaluations.

Various values for the cut-off frequency were considered, but the upper bound $\omega_{co} \leq 0.013$ rad/s was imposed throughout. This is not because the loads cannot provide service at higher frequencies, but because the response from uncertain loads should be attenuated near the crossover frequency $\omega_c = 0.05$ rad/s.

For the high-pass filter H_{HP}, a second-order Butterworth filter was used, with cut-off frequency of 0.0004 rad/s: $H_{HP}(s) = s^2/(s^2 + 0.0005657s + 1.6 \times 10^{-7})$. This was chosen based on the bandwidth constraints for pools, and QoS constraints for TCLs.

Further details are provided in the next subsection.

In the numerical experiments reported here, we have simulated the linear mean-field model of [64], and not the collection of loads. In all prior work it is found that the mean-field deterministic model reflects actual behavior very accurately, especially in a control setting. This is true even when only 100 loads are engaged [24]. A more detailed simulation taking into account transmission and perhaps distribution will be the subject of future work.

3.2 Load models and pre-filter design

The transfer function H in Figure 7.1 is obtained from a combination of resources, including demand dispatch. In prior work it has been shown that a randomized control architecture for each load leads to a tractable input-output model for the aggregate mean-field model [8]. Let us suppose that a load is modeled in a finite state space X with Markovian dynamics, X_t^i express the state of load i at time t and control targets a total number of N loads. Under certain assumptions, it is proven that the following limit holds:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{X_t^i = x\} = \mu_t(x), \quad x \in X.$$

One can therefore express the aggregate mean-field model by $\mu_{t+1} = \mu_t P_{\zeta_t}$. In this equation, μ_{t+1} is a row vector and P_{ζ_t} is the transition matrix constructed based on optimal control of loads supposing an input control signal ζ_t. The control is defined based on a desired aggregate power consumption of loads y_t expressed in power consumption at state x denoted by $U(x)$ by the following:

$$y_t = \sum_x \mu_t(x) U(x) \quad t \geq 0.$$

This mean-field model is nonlinear, but a linearized model can be obtained. Using Taylor series approximation, one can write $\mu_{t+1} \approx \mu_t P_0 + B^T \zeta_t$ where P_0 is the transition matrix of a control-free load and $B^T = \frac{d}{d\zeta} \pi P_\zeta$ evaluated at $\zeta = 0$ with π a unique equilibrium of a the control-free model. This linearized model worked well for purposes of estimation, control and performance evaluation [23, 24].

In the case of residential pools pumps, the focus of these three papers, the transfer function for the linear model has a strong resonance at a frequency corresponding to a period of 24 hours. The transfer function depends on the number of hours of cleaning per day, but the resonance is independent of this parameter. A second-order approximation was adopted,

$$G_{pl} = \frac{\omega_{pl}^2}{s^2 + 2\zeta_{pl}\omega_{pl}s + \omega_{pl}^2}, \tag{7.9}$$

with natural frequency $\omega_{pl} = 7.27 \times 10^{-5}$ rad/s; this corresponds to the 24-hour periodic behavior of pools.

The behavior of a typical TCL is similar to a pool filtration system. Take, for example, a typical residential refrigerator. Its nominal behavior is similar to a pool pump, with two exceptions. First, the behavior is roughly periodic with period much shorter than 24 hours. Second, in many cases the operating time is a smaller fraction of this period. We have obtained an input-output model using a technique similar to what is introduced in [8], and find that a collection of residential refrigerators with a 30 minute cycle-time admits a linear system model with resonance corresponding to this period. Accordingly, a second-order approximation for the aggregate of TCLs was used in these experiments,

$$G_{tcl} = \frac{\omega_{tcl}^2}{s^2 + 2\zeta_{tcl}\omega_{tcl}s + \omega_{tcl}^2}, \tag{7.10}$$

with natural frequency $\omega_{tcl} = 0.003$ rad/s, which approximately corresponds to the 30-minute cycle-time.

The damping ratios were set to $\zeta_{pl} = \zeta_{tcl} = 0.5$. Other values were tested with similar results.

To obtain the two transfer functions H_{pl} and H_{tcl} appearing in (7.8) requires one additional ingredient.

In prior work, it is shown that the bandwidth of ancillary service provided by loads can be extended through feedback. Even though the natural frequency for the linear model of [64] corresponds to 24 hours, a grid-level control solution extends the closed loop bandwidth by one decade, resulting in a closed-loop natural frequency of approximately 7.27×10^{-4} rad/s, and without resonance.

In the experiments described here, we do not allow communication from load to grid, so we cannot extend bandwidth using the approach of [64]. Instead, the bandwidth is set through the design of the pre-filter introduced in Section 2.2.

The local filters used at the pools and the TCLs are denoted M_{pl} and M_{tcl}, respectively. Following (7.2), we obtain the two actuator transfer functions,

$$H_{pl} = M_{pl}G_{pl} \qquad H_{tcl} = M_{tcl}G_{tcl} \tag{7.11}$$

The overall actuator transfer function of (7.8) is written,

$$H = H_a + H_b := P_a G_a + H_{LP}H_B \tag{7.12}$$

in which $H_B = [H_{pl} + \rho H_{HP}H_{tcl}]$ and

$$P_a = 1 - H_{LP}[1 - H_{HP}(1-\rho)] \tag{7.13}$$

When $\rho = 1$, the transfer function H is a convex combination of good and potentially "bad" actuators:

$$H = (1 - H_{LP})G_a + H_{LP}H_B$$

Two designs were considered for the local filter.

(i) *Lead design:* For $\tau > 0$ and $\alpha < 1$,

$$M_l = \Big(\frac{\tau s + 1}{\alpha\tau s + 1}\Big)^2 \tag{7.14}$$

(ii) *Inverse design:* For $\alpha < 1$,

$$M_l = (\alpha + G_l)^{-1} \tag{7.15}$$

The inverse design is intended to approximate the inverse of the transfer function for the aggregate load model. In very recent work, we have found that a better approximation can be obtained using concepts from robust control theory — this topic will be explored in detail in future work.

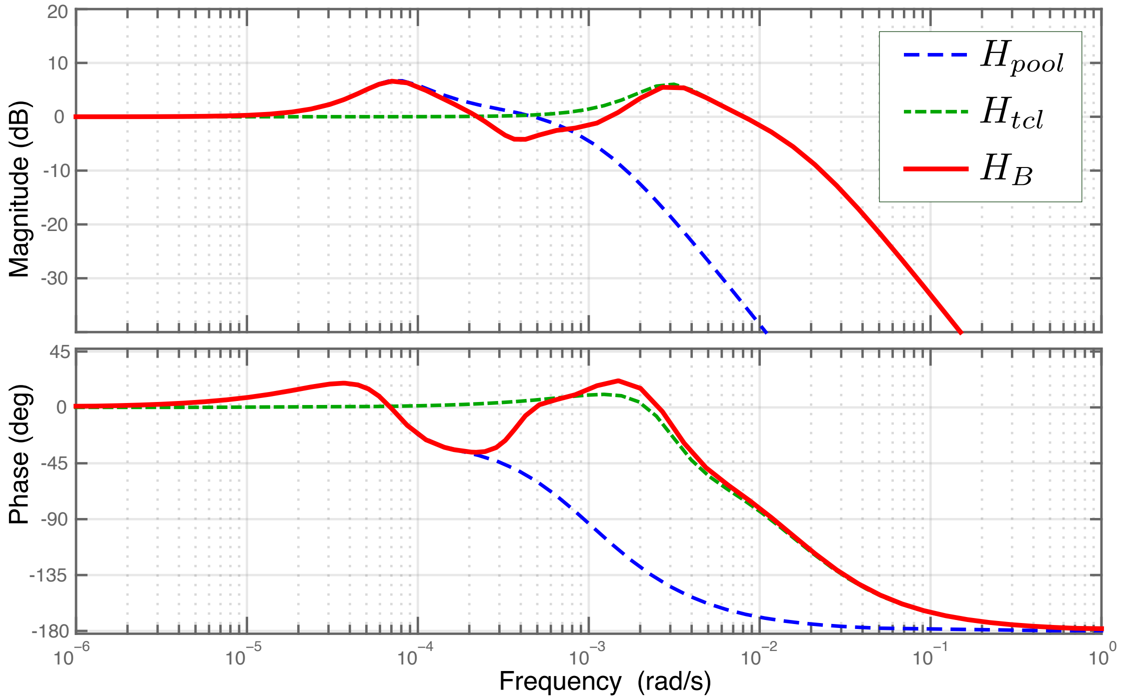


Figure 7.7 – H_B obtained using the lead design.

The lead design (7.14) consists of two lead compensators in series: this is to counter the −40 dB/decade slope after the resonance in the load transfer functions. The lead compensator is a high-pass filter, so the bandwidth of the loads is increased by some value depending on α and τ.

The lead design was applied to both pools and TCLs: we chose $\tau_{pl} = 15,000$ and $\tau_{tcl} = 350$, so that τ^{-1} is slightly smaller than the natural frequencies of the respective loads. In addition, we set $\alpha_{pl} = 1/15$ and $\alpha_{tcl} = 1/5$. Using these settings for the lead design, we achieve the required bandwidth extension for pools and TCLs.

The inverse design was also applied to both classes of loads. In the simulations that follow we only show results obtained when the inverse design was used to construct M_{tcl}; it was found that sensitivity of performance to the choice of the pre-filter M_{pl} is not great.

Figure 7.7 shows a Bode plot of H_B using the lead design with $\rho = 1$. The rapid phase decline at high frequencies motivates the cut-off frequency at $\omega_{co} = 0.003$ rad/s. We do not observe large phase lag using the inverse design for the TCLs, so we can take a larger value for ω_{co} in this case.

We will look more closely at the impact of these design choices in the following.

3.3 Transfer functions to model risk & performance

Recall that the loop transfer function is $L = G_c H G_p$. The corresponding *system sensitivity transfer function* is defined in control texts as

$$S = \frac{1}{1+L}, \tag{7.16}$$

and its maximum gain is denoted

$$M_S = \|S\|_\infty = \sup_\omega |S(j\omega)| \tag{7.17}$$

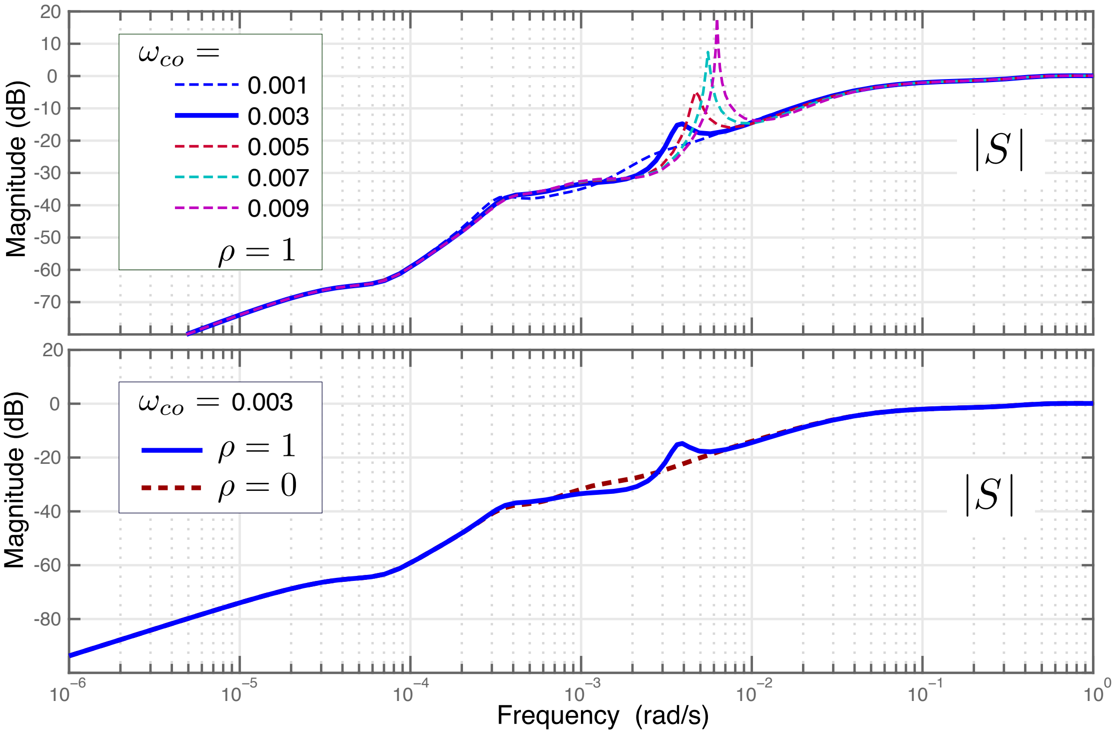


Figure 7.8 – Sensitivity using the lead design. The closed-loop system is not sensitive to the imperfect response from TCLs when $\omega_{co} = 0.005$ or smaller.

The terminology is motivated by the interpretation,

$$|S(j\omega)| = d(L(j\omega), -1)$$

where $d(L(j\omega), -1) = |L(j\omega) + 1|$. The minimum $\min_\omega |L(j\omega) + 1|$ is called the *vector margin* or the *stability margin*. If this is zero, then there is a closed loop pole on the imaginary axis, so the system is not stable. From the definitions, a small stability margin is equivalent to a large maximal sensitivity M_S.

Figure 7.8 shows the magnitude plot of the sensitivity function (7.16) for various values of ρ and ω_{co} using the lead design. The maximal sensitivity grows quickly with ω_{co} when $\rho = 1$. With $\omega_{co} = 0.003$ rad/s, we find that the sensitivity plots are similar for $\rho = 0$ or $\rho = 1$.

Sensitivity using the inverse design is lower for any $\omega_{co} \leq 0.013$ rad/s, and the value of ρ has little impact on sensitivity for this range of ω_{co}.

Recall from (7.5) that the transfer function from disturbance to output is $Y/D = G_p/(1 + L)$. Figure 7.9 shows Bode plots of this transfer function for the lead design: Using $\omega_{co} = 0.003$ rad/s, the system effectively suppresses disturbances whose frequency is below $\omega = 10^{-2}$. Performance degrades when using larger values of ω_{co}. For the inverse design, the transfer function Y/D is largely independent of ρ for any $\omega_{co} \leq 0.013$ rad/s.

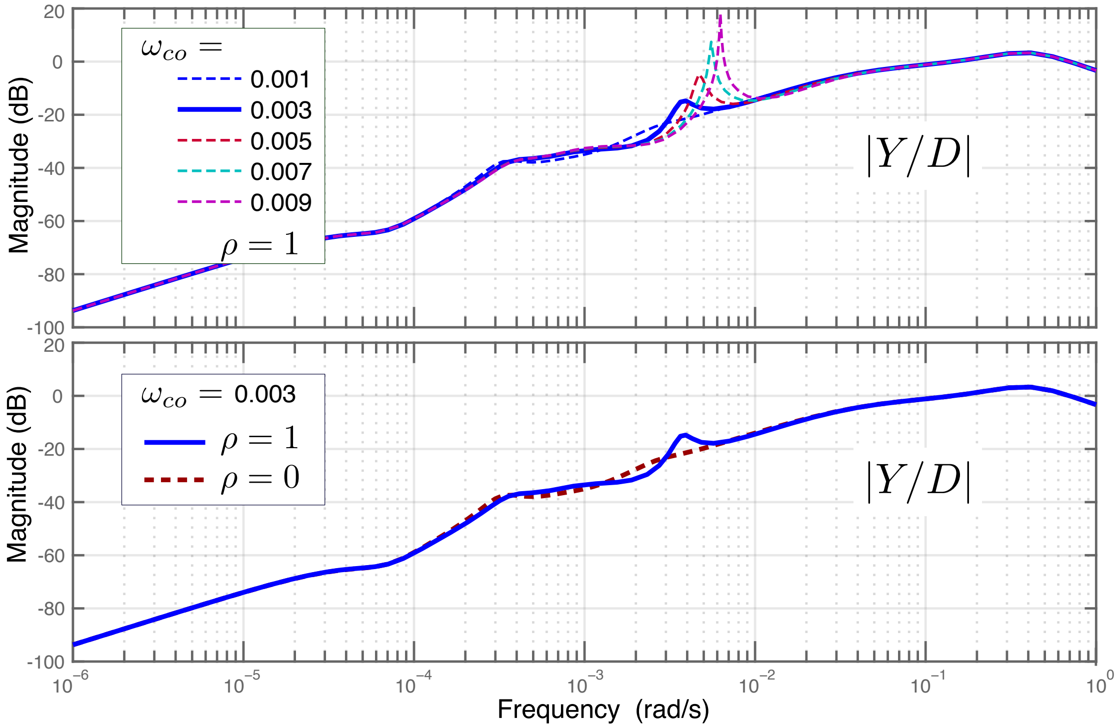


Figure 7.9 – Disturbance rejection using the lead design.

3.4 Mean-square cost

Let $U_a^{\mathbf{A}}(t)$ denote the signal supplied by the high-quality actuators. Based on the block diagram shown in Figure 7.1 and the definition of H in (7.8), we have

$$\frac{U_a^{\mathbf{A}}}{D} = -P_a \frac{L_a}{1+L}, \tag{7.18}$$

in which $L_a = G_c G_a G_p$. The proof of (7.18) is similar to the derivation of (7.6).

The maximal sensitivity is denoted

$$M_a = \left\| \frac{U_a^{\mathbf{A}}}{D} \right\|_\infty = \sup_\omega \left| \frac{U_a^{\mathbf{A}}}{D}(j\omega) \right| \tag{7.19}$$

While (7.19) is a common performance metric in the control literature, the maximizing frequency may have little to do with the frequencies encountered in operation.

The following L_2 cost is based on a statistical model of the disturbances entering the grid. We assume a steady-state setting in which the following *mean-square system cost* is well defined, and finite a.s.,

$$J^2 = \lim_{T\to\infty} \frac{1}{T} \int_0^T |U_a^{\mathbf{A}}(t)|^2 \, dt \tag{7.20}$$

This has a representation via spectral theory of stationary processes [16].

The power spectral density P_D for $\boldsymbol{D}$ can be estimated by fitting observed data to a linear system driven by white noise [16]. As done previously in [23], this approach was used to obtain an estimate

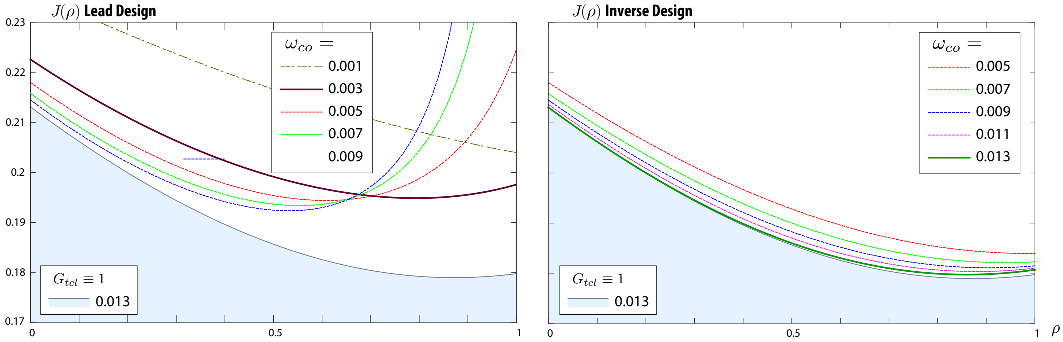


Figure 7.10 – Cost J as a function of ρ and ω_{co} for the two designs. The cost obtained using the inverse design is similar to what is obtained using an ideal actuator, $H_{tcl} \equiv 1$.

of P_D for the BPA balancing reserves. The resulting power spectral density for a stationary version of $\boldsymbol{U}_a^{\mathbf{A}}$ is given by

$$P_{U_a^{\mathbf{A}}}(\omega) = P_D(\omega)\left|\frac{U_a^{\mathbf{A}}}{D}(j\omega)\right|^2$$

where the transfer function is defined in (7.18). The mean-square system cost is expressed,

$$J^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} P_{U_a^{\mathbf{A}}}(\omega)\,d\omega\,. \tag{7.21}$$

Figure 7.10 shows two plots of J as a function of ρ for several values of ω_{co}. For comparison, in each plot, we include the cost plot for $\omega_{co} = 0.013$ rad/s and an "ideal TCL" in which $G_{tcl}(s) \equiv 1$.

In the lead design, the cost is near its minimum when $\rho = 1$ and $\omega_{co} = 0.003$ rad/s. In other words, with this lead design, we can minimize the system cost by including all the available TCL resources provided they service regulation signals with frequencies below 0.003 rad/s.

The bandwidth of TCL service can be increased significantly with more aggressive local control. For the inverse design, the cost is at its lowest for $\omega_{co} = 0.013$ rad/s. The performance nearly matches the ideal performance in this case.

3.5 Disturbance rejection

The disturbance-rejection performance of the system was evaluated using Simulink. Multiple experiments were conducted to evaluate performance using several different designs.

In each case, ρ was set to 1 in the definition of the actuator block H (see (7.8)). The remaining components of the block diagram shown in Figure 7.1 are as specified in this section.

In summary,

(i) When lead control is absent, the closed loop system is unstable for $\omega_{co} > 0.003$ rad/s, which is the natural frequency of G_{tcl}.

(ii) For the lead design, disturbance rejection is very good for $\omega_{co} \leq 0.005$ rad/s; the system is unstable for $\omega_{co} \geq 0.009$ rad/s.

Figure 7.6a shows results for $\omega_{co} = 0.007$ rad/s. The oscillation is due to a strong resonance in the closed loop system — this is caused by phase lag in H_{tcl}.

(iii) Disturbance rejection is improved using the inverse design for the TCLs. Results obtained with $\omega_{co} = 0.013$ rad/s are shown in Figure 7.6b. The disturbance rejection observed here is the best among all experiments.

4 Conclusion

In this chapter, we investigate the possibility of a decentralized control scheme based on randomised local control. We study the impact of global tracking performance on system wide costs. We show that poor performance at low frequency ranges incur little cost on the system. However, when combining poor performance with performing resources, the system may suffer high costs. Indeed, in the case of heterogeneity of resources, performing ones are forced to provide greater service to the grid to compensate for the poor performance of other resources.

In addition, we show that with appropriate filtering and local control, excellent performance can be seen in low-to-mid frequency ranges. These results are very promising in terms of preserving consumers privacy even if it may incur some installation and configuration costs.

Chapter 8

Conclusion

In this thesis, we propose and analyze advanced DR solutions that can take advantage of heterogeneous appliances' flexibility in order to reach required load shape objectives. Indeed, flexibility provided by resources participating to DR is a major asset in maintaining the balance between generation and consumption particularly during periods of scarce power availability. In Chapter 2, we give an overview of approaches that enable to meet such objectives and discuss the need for joint optimization of heterogeneous resources.

Taking decisions on the level of a generic appliance is challenging: it should depend on the energy profile of the appliance and the impact of altering this profile on user's perceived quality of experience. In Chapter 3, we build on top of existing classification efforts and define a new generic framework allowing to describe any appliance's power profile flexibility. Hence, this framework provides the means to build a federated ecosystem between heterogeneous resources. For this, we approximate appliances desired profile through the definition of utility functions. These functions are utilized to derive the control schemes proposed in this thesis. We show that the proposed model for utility functions is rich and precise enough to represent appliances serving very different purposes. As for the user, she can attribute to each appliance in her home a criticality level (e.g., vital and comfort) and a priority (between appliances operations of the same criticality). This allows different users to have different preferences for the same appliance. We also propose a way to deal with users having different quality of service requirements.

Following chapters evaluate the performance of mechanisms based on the proposed utility model. These mechanisms act at different degrees of granularity for information and knowledge of network constraints (centralized, partially distributed and distributed) and suppose different decision making timescales (i.e., ahead-of-time or real-time). Proposed families of schemes are presented in Table 8.1.

Chapter 4 focuses on a fully centralized control architecture. Proposed solutions are formulated so that utility perceived by end users is maximized under system constraints. This allows us to present through simple examples how control schemes can be developed through the framework proposed in Chapter 3. When a decision is made ahead-of-time, proposed centralized control provides the globally optimal solution utilized as a benchmark to assess the performance of all the other proposed solutions. We introduce two fairness criteria: one based on lexicographic ordering of utility functions, and one improves the worst case scenario for homes. Through numerical analysis, we show the advantage of the first fairness criterion by studying a two-level utility model and single level one. In practice, it may be difficult to reach the global optimum since resolution time exponentially increase with the number of appliances controlled. This makes an exact resolution hard in a realistic setting. Thus,

Schemes	Performance	Complexity	Privacy	Chapter
Centralized	+++	-(exact) → +++ (heuristic)	-	4
Partially distributed with one-way communication	+	++	+++	5 & 7
Partially distributed with two-way communication	++	++	+	5
Distributed	+++	++	++	6

Table 8.1 – Proposed schemes summary

we propose approximation methods based on greedy algorithms. Numerical results obtained show the high importance of defining appliances utility functions model in view of decision time-scale and optimization horizon. Indeed, decisions need to consider the real value of operating an appliance. Regardless of the advantages provided by approximation methods in terms of improved scalability, a centralized solution causes privacy issues. The next chapters aim at proposing architectures where decisions are distributed. Indeed, proposed schemes target to reach the performances of a centralized solution while improving scalability and providing privacy guaranties.

In Chapter 5, we consider a more realistic control architecture based on “divide and conquer”: an aggregator that controls homes consumptions, and home decision entities that schedule appliances locally based on aggregator's instructions. Two schemes based on one-way communication are proposed: one inspired from round-robin like schemes traditionally used in power systems to deal with lack of generation capacity, and one that divides capacity proportionally between homes. These schemes rely on static home information and thus may render allocations that are inefficient. We show that while both schemes fail to reach the performance of a centralized solution, proportional split of available capacity provides a better quality of experience than the scheme inspired by the solution adopted currently to deal with lack of capacity.

To overcome the limitations of one-way schemes that rely on static information, we propose a scheme inspired from gradient decomposition that considers a two-way communication. In this case, privacy is enforced by disclosing only aggregated information as a feedback from homes. We show that this scheme provides better performance in terms of perceived utility by end users compared to one-way schemes in situations where capacity is very low. While schemes proposed in this chapter meet privacy requirements, performance can be further improved. This goal is addressed in Chapter 6 and its limitations are evaluated in Chapter 7.

In Chapter 6, we propose distributed control solutions inspired by peer-to-peer systems. Supposing an initial capacity distribution, homes can coordinate in order to improve the efficiency of their allocations. This is done by allowing home to dynamically contact other homes to test the possibility of a power exchange. Schemes are discussed particularly in view of users selfishness and the number of coordinating homes. Through numerical analysis, we show the high performance of distributed schemes that provide near-optimal allocations in most cases. We also discuss the fallout of selfishness of houses on the convergence speed and the quality of the solution especially when homes of heterogeneous needs are supposed. In addition, we show that a three homes coordination has better performance than that a of coordination between two homes.

While previous chapters focus on strict control, Chapter 7 analyzes a system where appliances can decide how they react to control signals. In this chapter, we analyze the effect of the accuracy of resources' response on overall system cost. Cost is defined in terms of required expensive high performing resources to maintain grid stability. We show that poor performing resources can incur high system cost when they co-exist with high performing resources. We also show that with proper design and supposing intelligence at loads, a decentralized control scheme can be proposed allowing to have excellent performance in certain cases without requiring any feedback.

In this thesis, we show the feasibility of a generic framework that can integrate heterogeneous resources to provide Demand Response services to the grid. Our results, summarized in Table 8.1, can be seen as follows: centralized schemes are the best for performance while their computational complexity can be addressed using carefully designed greedy heuristics. Privacy concerns raised by full information disclosure may prevent their adoption. Privacy can be fully or partially addressed considering one-way or two-way communication schemes at the cost of a lower performance. Finally, for our use cases, it seems that distributed schemes can provide near-optimal performance while addressing privacy by eliminating the need for central coordination.

1 Future work

The contributions of this thesis are a first stepping stone towards designing efficient DR solutions capable of integrating heterogeneous resources. We briefly outline possible future research directions:

Local physical constraints
In the present thesis, we focus on evaluating the impact on performance of several types of control schemes supposing simple grid side requirements. These requirements are formalized in terms of the desired shape of aggregate consumption curve. However, it might be desired to take advantage of loads location in order to provide more effective response depending on local constraints of the physical grid. For instance, when local generations are deployed on the distribution system, it is desired to consume their generation locally to minimize line losses.

DR value and incentives
We proposed DR solutions that are capable of taking required decisions to reach control objectives while taking into account users quality of experience. High performance of the control will increase its attractiveness. However, additional incentives need to be provided to users to make them want to participate to DR. This requires the development of key performance indicators that allow to value the service actually provided by users. Today, this remains an open question. For example, it is very challenging to predict what would have been the behavior of users if no control is applied, in order to quantify how much flexibility they offered.

Demand Response under uncertainty
This work is the first step towards the design of a system that can take into account parameters uncertainty. We provided intuition required when developing efficient algorithms to solve a deterministic optimization problem targeting a set of controlled appliances. These algorithms can be used to solve stochastic optimization problems. For instance, techniques like scenario decomposition allow to derive deterministic equivalent problems from possible realizations of uncertain parameters. Dealing with uncertainty can also be done by re-triggering control decisions computation when parameters deviate from their expected value.

Manipulation and truthfulness

In this thesis, we suppose truthful users and correct data acquisition. Means to detect users behavior that try to game the system is needed. In our proposed model, several parameters values are needed to take adequate control decisions. Some of these values can be provided by users and others are collected from sensors. If sensors' fine grained data is reliable, dealing with truthfulness needs to be addressed by designing control algorithms that can verify the sincerity of users without inducing privacy issues. Additional constraints can also be considered to control the maximum amount of energy consumed during the optimization period.

Acronyms

ACE Area Control Error.

AMI Advanced Metering Infrastructure.

A/S Ancillary Services.

BPA Bonneville Power Administration.

BRD Balancing Reserves Deployed.

CAISO California Independent System Operator.

CPP Critical Peak Pricing.

DHT Distributed Hash Table.

DLC Direct Load Control.

DR Demand Response.

DSM Demand Side Management.

DSO Distribution System Operator.

ETPSG European Technology Platform Smart Grid.

EV Electric Vehicle.

FERC Federal Energy Regulatory Commission.

IBR Inclining Block Rate.

ICT Information and Communication Technologies.

ID Identifier.

IESO Independent Electricity System Operator

ISO Independent System Operator

IoT Internet of Things.

microCHP Micro Combined Heat and Power.

MILP Mixed Integer Linear Problem.

NILM Non-Intrusive Load Monitoring.

NERC North American Electric Reliability Corporation.

NREL National Renewable Energy Laboratory.

NYISO New York Independent System Operator.

PJM Pennsylvania-New Jersey-Maryland Interconnection.

PSD Power Spectral Density.

PTR Peak-Time Rebates.

PV PhotoVoltaics.

QoS Quality of Service.

RSO Regional System Operator

RTO Regional Transmission Organization.

RTP Real-Time Pricing.

SEP Smart Energy Profile.

SGAM Smart Grid Architecture Model.

TCL Thermostatically Controlled Load.

TOU Time Of Use.

TSO Transmission System Operator.

USA United States of America.

Bibliography

[1] Alessandro Agnetis, Gabriella Dellino, Gianluca De Pascale, Giacomo Innocenti, Marco Pranzo, and Antonio Vicino. Optimization models for consumer flexibility aggregation in smart grids: The address approach. *Smart Grid Modeling and Simulation (SGMS), 2011 IEEE First International Workshop on*, pages 96–101, 2011.

[2] Alessandro Agnetis, Gabriella Dellino, Paolo Detti, Giacomo Innocenti, Gianluca De Pascale, and Antonio Vicino. Appliance operation scheduling for electricity consumption optimization. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 5899–5904. IEEE, 2011.

[3] Italo Atzeni, Luis G Ordóñez, Gesualdo Scutari, Daniel P Palomar, and Javier R Fonollosa. Demand-side management via distributed energy generation and storage optimization. *Smart Grid, IEEE Transactions on*, 4(2):866–876, 2013.

[4] Yossi Azar, Andrei Z. Broder, Anna R. Karlin, and Eli Upfal. Balanced allocations. *SIAM J. Comput.*, 29(1):180–200, September 1999.

[5] Adil M Bagirov, Bülent Karasözen, and Meral Sezer. Discrete gradient method: derivative-free method for nonsmooth optimization. *Journal of Optimization Theory and Applications*, 137(2):317–334, 2008.

[6] Tanuja Bapat, Neha Sengupta, Sunil Kumar Ghai, Vijay Arya, Yedendra Babu Shrinivasan, and Deva Seetharam. User-sensitive scheduling of home appliances. *Proceedings of the 2nd ACM SIGCOMM workshop on Green networking*, 2011.

[7] Galen Barbose, Charles Goldman, and Bernie Neenan. A survey of utility experience with real time pricing. *Lawrence Berkeley National Laboratory*, 2004.

[8] Prabir Barooah, Ana Bušić, and Sean Meyn. Spectral decomposition of demand-side flexibility for reliable ancillary services in a smart grid. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, pages 2700–2709. IEEE, 2015.

[9] Kari Evans Bennett. Status of MISO's regulation market since implementation of the performance-based revisions as required by Order No. 755. Midcontinent Independent System Operator, Inc. – Informational Report. Technical report, Midcontinent Independent System Operator, 2013. Report prepared for the FERC. Docket # ER12-1664, Midwest ISO, Senior Corporate Counsel.

[10] Ontario Energy Board. Demand-side management and demand response in the ontario energy sectors. Technical report, Ministry of Energy, 2004. http://www.ontarioenergyboard.ca/documents/cases/RP-2003-0144/pressrelease_report_finalwithappendices_030304.pdf.

[11] Severin Borenstein. The long-run efficiency of real-time electricity pricing. *The Energy Journal*, pages 93–116, 2005.

[12] Stephen Boyd, Lin Xiao, and Almir Mutapcic. Subgradient methods. *lecture notes of EE392o, Stanford University, Autumn Quarter*, 2004:2004–2005, 2003.

[13] Bonneville Power Administration balancing authority. http://transmission.bpa.gov/business/operations/default.aspx.

[14] Kevin Martin Brokish. *Adaptive load control of microgrids with non-dispatchable generation*. PhD thesis, Massachusetts Institute of Technology, 2009.

[15] Alec Brooks, Ed Lu, Dan Reicher, Charles Spirakis, and Bill Weihl. Demand dispatch. *Power and Energy Magazine, IEEE*, 8(3):20–29, 2010.

[16] Peter E Caines. *Linear stochastic systems*. John Wiley & Sons, Inc., 1987.

[17] Peter Cappers, Charles Goldman, and David Kathan. Demand response in us electricity markets: Empirical evidence. *Energy*, 35(4):1526–1535, 2010.

[18] François Caron, Fabienne Cardot, Maurice Lévy-Leboyer, Michel Banal, Henri Morsel, et al. *Histoire générale de l'électricité en France*. Fayard, 1996.

[19] Tommy Carpenter, Sahil Singla, Parsiad Azimzadeh, and Srinivasan Keshav. The impact of electricity pricing schemes on storage adoption in ontario. In *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet*, page 18. ACM, 2012.

[20] Douglas W Caves and Laurits R Christensen. Econometric analysis of residential time-of-use electricity pricing experiments. *Journal of Econometrics*, 14(3):287–306, 1980.

[21] Harold R Chamorro, Mehrdad Ghandhari, and Robert Eriksson. Influence of the increasing non-synchronous generation on small signal stability. In *PES General Meeting| Conference & Exposition, 2014 IEEE*, pages 1–5. IEEE, 2014.

[22] Hector Chavez, Ross Baldick, and Sandip Sharma. Regulation adequacy analysis under high wind penetration scenarios in ERCOT nodal. *Sustainable Energy, IEEE Transactions on*, 3(4):743–750, 2012.

[23] Yue Chen, Ana Bušić, and Sean Meyn. Individual risk in mean field control with application to automated demand response. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 6425–6432. IEEE, 2014.

[24] Yue Chen, Ana Bušić, and Sean Meyn. State estimation and Mean-Field Control with application to demand dispatch. *arXiv preprint arXiv:1504.00088*, 2015.

[25] Zhi Chen and Lei Wu. Residential appliance dr energy management with electric privacy protection by online stochastic optimization. *Smart Grid, IEEE Transactions on*, 4(4):1861–1869, 2013.

[26] Zhi Chen, Lei Wu, and Yong Fu. Real-time price-based demand response management for residential appliances via stochastic optimization and robust optimization. *Smart grid, IEEE transactions on*, 3(4):1822–1831, 2012.

[27] Pilong Chris. Pjm manual 12: Balancing operations, 2012.

[28] Angela S Chuang and Clark W Gellings. Demand-side integration in a restructured electric power industry. *CIGRE International council on large electric systems, number Paper C6-105, Session*, 2008.

[29] Toby Considine, William Cox, and Edward G Cazalet. Understanding microgrids as the essential architecture of smart energy. In *Grid Inerop Forum*, 2012.

[30] Ruilong Deng, Rongxing Lu, Gaoxi Xiao, and Jiming Chen. Fast distributed demand response with spatially-and temporally-coupled constraints in smart grid. *Industrial informatics Grid, IEEE Transactions on*, 2015.

[31] Pengwei Du and Ning Lu. Appliance commitment for household load scheduling. *Smart Grid, IEEE Transactions on*, 2(2):411–419, 2011.

[32] Sustainable Development Task Force. Regulatory and market aspects of demand-side flexibility. Technical report, Council of European Energy Regulators, 2013. http://www.ceer.eu/portal/page/portal/EER_HOME/EER_CONSULT/CLOSED%20PUBLIC%20CONSULTATIONS/ELECTRICITY/Demand-side_flexibility/CD/C13-PC-71_C13-SDE-38-03_SF_PC_2013-11-04.pdf.

[33] Michael R. Garey and Johnson David S. *Computers and intractibility: A guide to the Theory of NP-completeness*. W.H. Freeman and Company, 1979.

[34] Nikolaos Gatsis and Georgios B Giannakis. Cooperative multi-residence demand response scheduling. *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*, pages 1–6, 2011.

[35] Nikolaos Gatsis and Georgios B Giannakis. Residential load control: Distributed scheduling and convergence with lost ami messages. *Smart Grid, IEEE Transactions on*, 3(2):770–786, 2012.

[36] Jean-Michel Glachant and Yannick Perez. L'analyse économique appliquée à la problématique des effacements diffus. *Revue de l'énergie*, 597:312–321, 2010.

[37] Junqiao Han and Mary Ann Piette. Solutions for summer electric power shortages: Demand response and its applications in air conditioning and refrigerating systems. *Refrigeration, Air Conditioning and Electric Power Machinery*, 29(1):1–4, 2008.

[38] He Hao, Timothy Middelkoop, Prabir Barooah, and Sean Meyn. How demand response from commercial buildings will provide the regulation needs of the grid. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 1908–1913. IEEE, 2012.

[39] Barry Hayes, Ignacio Hernando-Gil, Adam Collin, Gareth Harrison, and Sasa Djokic. Optimal power flow for maximizing network benefits from demand-side management. *Power Systems, IEEE Transactions on*, 29(4):1739–1747, 2014.

[40] Michael Held, Philip Wolfe, and Harlan P Crowder. Validation of subgradient optimization. *Mathematical programming*, 6(1):62–88, 1974.

[41] John Holton. *Strategy Guideline: High Performance Residential Lighting*. US Department of Energy, Energy Efficiency & Renewable Energy, Building Technologies Program, 2012.

[42] Zheng Hu, Jin-ho Kim, Jianhui Wang, and John Byrne. Review of dynamic pricing programs in the us and europe: Status quo and policy recommendations. *Renewable and Sustainable Energy Reviews*, 42:743–751, 2015.

[43] Kun-Yuan Huang and Yann-Chang Huang. Integrating direct load control with interruptible load management to provide instantaneous reserves for ancillary services. *Power Systems, IEEE Transactions on*, 19(3):1626–1634, 2004.

[44] IBM Global Business Services and eMeter Strategic. Ontario energy board smart price pilot final report. Technical report, Ontario Energy Board, 2007. http://www.ontarioenergyboard.ca/documents/cases/EB-2004-0205/smartpricepilot/OSPP%20Final%20Report%20-%20Final070726.pdf.

[45] IBM Ilog CPLEX optimizer. http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/.

[46] Ali Ipakchi and Farrokh Albuyeh. Grid of the future. *Power and Energy Magazine, IEEE*, 7(2):52–62, 2009.

[47] Carlee Joe-Wong, Soumya Sen, Sangtae Ha, and Mung Chiang. Optimized day-ahead pricing for smart grids with device-specific scheduling flexibility. *Selected Areas in Communications, IEEE Journal on*, 30(6):1075–1085, 2012.

[48] Gopinath Karmakar, Ashutosh Kabra, and Krithi Ramamritham. Coordinated scheduling of thermostatically controlled realtime systems under peak power constraint. *IEEE 19th RealTime and Embedded Technology and Applications Symposium (RTAS)*, 2013.

[49] Brendan J Kirby. *Frequency regulation basics and trends*. United States. Department of Energy, 2005.

[50] Daniel Kofman and Catherine Rosenberg. New service models for electricity distribution during energy crises. Technical report, Telecom ParisTech, 2012. http://service.tsi.telecom-paristech.fr/cgi-bin/valipub_download.cgi?dId=265.

[51] George Koutitas. Control of flexible smart devices in the smart grid. *Smart Grid, IEEE Transactions on*, 3(3):1333–1343, 2012.

[52] Prabha Kundur, Neal J Balu, and Mark G Lauby. *Power system stability and control*, volume 7. McGraw-hill New York, 1994.

[53] HY Lam, GSK Fung, and WK Lee. A novel method to construct taxonomy electrical appliances based on load signatures. *Consumer Electronics, IEEE Transactions on*, 53(2):653–660, 2007.

[54] Jean-Yves Le Boudec and Dan-Cristian Tomozei. Demand response using service curves. In *Innovative Smart Grid Technologies (ISGT Europe), 2011 2nd IEEE PES International Conference and Exhibition on*, pages 1–8. IEEE, 2011.

[55] Na Li, Lijun Chen, and Steven H Low. Optimal demand response based on utility maximization in power networks. *IEEE Power and Energy Society General Meeting*, 2011.

[56] Na Li, Lingwen Gan, Lijun Chen, and Steven H Low. An optimization-based demand response in radial distribution networks. In *Globecom Workshops (GC Wkshps), 2012 IEEE*, pages 1474–1479. IEEE, 2012.

[57] Jian Liang, Simon KK Ng, Gail Kendall, and John WM Cheng. Load signature study—part i: Basic concept, structure, and methodology. *Power Delivery, IEEE Transactions on*, 25(2):551–560, 2010.

[58] Yong Liang, Long He, Xinyu Cao, and Zuo-Jun Shen. Stochastic control for smart grid users with flexible demand. *Smart Grid, IEEE Transactions on*, 4(4):2296–2308, 2013.

[59] Yashen Lin, Prabir Barooah, Sean Meyn, and Timothy Middelkoop. Experimental evaluation of frequency regulation from commercial building hvac systems. *Smart Grid, IEEE Transactions on*, 6(2):776–783, 2015.

[60] Yi Liu, Chau Yuen, Shisheng Huang, Naveed Ul Hassan, Xiumin Wang, and Shengli Xie. Peak-to-average ratio constrained demand-side management with consumer's preference in residential smart grid. *Selected Topics in Signal Processing, IEEE Journal of*, 8(6):1084–1097, 2014.

[61] Eng Keong Lua, Jon Crowcroft, Marcelo Pias, Ritu Sharma, and Sharon Lim. A survey and comparison of peer-to-peer overlay network schemes. *Communications Surveys & Tutorials, IEEE*, 7(2):72–93, 2005.

[62] Johanna L Mathieu, Maryam Kamgarpour, John Lygeros, and Duncan S Callaway. Energy arbitrage with thermostatically controlled loads. *European Control Conference (ECC)*, 2013.

[63] Petar Maymounkov and David Mazieres. Kademlia: A peer-to-peer information system based on the xor metric. *Peer-to-Peer Systems*, pages 53–65, 2002.

[64] Sean Meyn, Prabir Barooah, Ana Bušić, Yue Chen, and Jordan Ehren. Ancillary service to the grid using intelligent deferrable loads. *arXiv preprint arXiv:1402.4600*, 2014.

[65] NW Miller, M Shao, S Pajic, and R D'Aquila. Western wind and solar integration study phase 3–frequency response and transient stability (report and executive summary). Technical report, National Renewable Energy Laboratory (NREL), Golden, CO., 2014.

[66] Aditya Mishra, David Irwin, Prashant Shenoy, Jim Kurose, and Ting Zhu. Smartcharge: cutting the electricity bill in smart homes with energy storage. In *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet*, page 29. ACM, 2012.

[67] Amir-Hamed Mohsenian-Rad and Alberto Leon-Garcia. Optimal residential load control with price prediction in real-time electricity pricing environments. *Smart Grid, IEEE Transactions on*, 1(2):120–133, 2010.

[68] Amir-Hamed Mohsenian-Rad, Vincent WS Wong, Juri Jatskevich, and Robert Schober. Optimal and autonomous incentive-based energy consumption scheduling algorithm for smart grid. *Innovative Smart Grid Technologies (ISGT)*, pages 1–6, 2010.

[69] Amir-Hamed Mohsenian-Rad, Vincent WS Wong, Juri Jatskevich, Robert Schober, and Alberto Leon-Garcia. Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. *Smart Grid, IEEE Transactions on*, 1(3):320–331, 2010.

[70] Albert Molderink, Vincent Bakker, Maurice GC Bosman, Johann L Hurink, and Gerard JM Smit. Management and control of domestic smart grid technology. *Smart Grid, IEEE Transactions on*, 1(2):109–119, 2010.

[71] Benyamin Moradzadeh and Kevin Tomsovic. Two-stage residential energy management considering network operational constraints. *Smart Grid, IEEE Transactions on*, 4(4):2339–2346, 2013.

[72] Ashutosh Nayyar, Josh Taylor, Anand Subramanian, Kameshwar Poolla, and Pravin Varaiya. Aggregate flexibility of a collection of loads π. *IEEE Conference on Decision and Control*, 2013.

[73] U.S. Department of Energy. Benefits of demand response in electricity markets and recommendations for achieving them. Technical report, United States Congress, 2006. http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/DOE_Benefits_of_Demand_Response_in_Electricity_Markets_and_Recommendations_for_Achieving_Them_Report_to_Congress.pdf.

[74] Michael Angelo A Pedrasa, Ted D Spooner, and Iain F MacGill. Coordinated scheduling of residential distributed energy resources to optimize smart home energy services. *Smart Grid, IEEE Transactions on*, 1(2):134–143, 2010.

[75] Cristian Perfumo, Ernesto Kofman, Julio H. Braslavsky, and John K. Ward. Load management: Model-based control of aggregate power for populations of thermostatically controlled loads. *Energy Conversion and Management*, 55(0):36 – 48, 2012.

[76] Mansoureh Peydayesh and Ross Baldick. The effects of very fast response to frequency fluctuation. In *Proceedings of the 31st USAEE/IAEE North American Conference*. International Association for Energy Economics, 2012.

[77] Mary Ann Piette. Linking continuous energy management and open automated demand response. *Lawrence Berkeley National Laboratory*, 2009.

[78] Peeter Pikk and Marko Viiding. The dangers of marginal cost based electricity pricing. *Baltic Journal of Economics*, pages 49–62, 2013.

[79] Michal Pioro and Deepankar Medhi. *Routing, Flow and Capacity Design in Communication and Computer Networks*. Morgan Kaufmann Publishers, 2004.

[80] Energy Pool. CEER annual conference unlocking energy market flexibility and demand side response. http://www.ceer.eu/portal/page/portal/EER_HOME/EER_WORKSHOP/CEER-ERGEG%20EVENTS/CEER_Conferences/CEER_CONFERENCE_2015/Presentations/Energy_Pool_Presentation_final.pdf, 2015.

[81] Seyyed Ali Pourmousavi and M. Hashem Nehrir. Introducing dynamic demand response in the LFC model. *Power Systems, IEEE Transactions on*, 29(4):1562–1572, July 2014.

[82] Navid Rahbari-Asr and Mo-Yuen Chow. Cooperative distributed demand management for community charging of phev/pevs based on kkt conditions and consensus networks. *Industrial Informatics, IEEE Transactions on*, 10(3):1907–1916, 2014.

[83] Yann G Rebours, Daniel S Kirschen, Marc Trotignon, and Sébastien Rossignol. A survey of frequency and voltage control ancillary services—part i: Technical features. *Power Systems, IEEE Transactions on*, 22(1):350–357, 2007.

[84] Mardavij Roozbehani, Munther Dahleh, Sanjoy K Mitter, et al. Volatility of power grids under real-time pricing. *Power Systems, IEEE Transactions on*, 27(4):1926–1940, 2012.

[85] Amir Safdarian, Mahmud Fotuhi-Firuzabad, and Matti Lehtonen. A distributed algorithm for managing residential demand response in smart grids. *Industrial Informatics, IEEE Transactions on*, 10(4):2385–2393, 2014.

[86] Pedram Samadi, Amir-Hamed Mohsenian-Rad, Robert Schober, Vincent WS Wong, and Juri Jatskevich. Optimal real-time pricing algorithm based on utility maximization for smart grid. *First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 415–420, 2010.

[87] Borhan M Sanandaji, He Hao, Kameshwar Poolla, and Tyrone L Vincent. Improved battery models of an aggregation of thermostatically controlled loads for frequency regulation. *arXiv preprint arXiv:1310.1693*, 2013.

[88] Fred C Schweppe, Richard D Tabors, James L Kirtley Jr, Hugh R Outhred, Frederick H Pickel, and Alan J Cox. Homeostatic utility control. *Power Apparatus and Systems, IEEE Transactions on*, pages 1151–1163, 1980.

[89] Wenbo Shi, Na Li, Xiaorong Xie, Chi-Cheng Chu, and Rajit Gadh. Optimal residential demand response in distribution networks. *IEEE journal on selected areas in communications*, 2014.

[90] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.

[91] Bethany Speer, Mackay Miller, Walter Shaffer, Leyla Gueran, Albrecht Reuter, and Bonnie and Jang. The role of smart grid in integrating renewable energy. Technical report, National Renewable Energy Laboratory, 03 2015.

[92] Pirathayini Srikantha, Srinivasan Keshav, and Catherine Rosenberg. Distributed control for reducing carbon footprint in the residential sector. *IEEE smartgridcomm*, 2012.

[93] Pirathayini Srikantha, Catherine Rosenberg, and Srinivasan Keshav. An analysis of peak demand reductions due to elasticity of domestic appliances. *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet*, page 28, 2012.

[94] Pirathayini Srikantha, Catherine Rosenberg, and Srinivasan Keshav. An analysis of peak demand reductions due to elasticity of domestic appliances. *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet*, page 28, 2012.

[95] Rainer Stamminger and Rheinische Friedrich. Synergy potential of smart appliances. Technical report, Smart-A project, 2008. http://www.smart-a.org/WP2_D_2_3_Synergy_Potential_of_Smart_Appliances.pdf.

[96] Doug Vine and Timothy Juliant. Climate solutions: the role of nuclear power. *Center for Climate and Energy Solutions*, 2014.

[97] Cynthujah Vivekananthan, Yateendra Mishra, Gerard Ledwich, and Fangxing Li. Demand response for residential appliances via customer reward scheme. *IEEE Trans. Smart Grid*, 5(2):809–820, 2014.

[98] Sunil K Vuppala, Kumar Padmanabh, Sumit Kumar Bose, and Sanjoy Paul. Incorporating fairness within demand response programs in smart grid. In *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, pages 1–9. IEEE, 2011.

[99] Rahul Walawalkar, Stephen Fernands, Netra Thakur, and Konda Reddy Chevva. Evolution and current status of demand response (dr) in electricity markets: Insights from pjm and nyiso. *Energy*, 35(4):1553–1560, 2010.

[100] David P Williamson and David B Shmoys. *The design of approximation algorithms*. Cambridge University Press, 2011.

[101] Ying Xiao, Qianli Su, Stu Bresler, S Frederick, Rebecca Carroll, Jeffrey R Schmitt, and Michael Olaleye. Performance-based regulation model in PJM wholesale markets. In *PES General Meeting| Conference & Exposition, 2014 IEEE*, pages 1–5. IEEE, 2014.

[102] Ahmed Zoha, Alexander Gluhak, Muhammad Ali Imran, and Sutharshan Rajasegarar. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12(12):16838–16866, 2012.

Gestion Active de la Demande Basée sur l'Habitat Connecté

Rim Kaddah

RESUME :

L'Internet des Objets (IdO) et le déploiement des équipements connectés permettent la mise en place de solutions de Gestion Active de la Demande(GAD) avancées. En effet, il devient possible d'avoir plus de visibilité et un contrôle fin sur différents équipements qui consomment, stockent ou produisent de l'énergie dans une maison. Dans cette thèse, nous considérons des solutions ayant la capacité de produire des décisions de contrôle direct à différents niveaux de granularité en fonction des variables mesurées dans les habitats. Le contrôle est basé sur une optimisation d'utilité perçue. Des fonctions utilité sont définies à travers une approche générique qui considère la flexibilité de la charge et l'impact des décisions de contrôle sur les utilisateurs. L'approche proposée n'impose pas de restrictions sur le type des équipements contrôlés ni sur la granularité des décisions de contrôle. Ceci permet un contrôle joint d'équipements hétérogènes. Nous considérons trois types d'architectures de contrôle à savoir: des solutions centralisées, partiellement distribuées et entièrement distribuées. Ces architectures diffèrent dans la distribution de la prise de décision entre les entités impliquées dans le contrôle et les données qui sont mis à disposition de ces entités. L'analyse numérique montre les compromis des solutions proposées du point de vue de la performance, de l'extensibilité et de la complexité.

MOTS-CLEFS: Gestion active de la demande, contrôle direct des équipements, réseau électrique intelligent, Internet des Objets, recherche opérationnelle

ABSTRACT:

The Internet of Things (IoT) paradigm brings an opportunity for advanced Demand Response (DR) solutions. Indeed, it enables visibility and control on the various appliances that may consume, store or generate energy within a home. In this thesis, we consider solutions having the capability to produce direct control decisions at different granularities based on variables measured at homes. Control schemes are driven by an optimization based on utility functions. These functions are defined based on a generic approach that considers load's flexibility and the impact of control decisions on users. The proposed approach does not impose any restrictions on the type of controlled appliances nor on the granularity of control decisions. This enables joint control of heterogeneous loads. We consider three types of control architectures, namely centralized, partially distributed and fully distributed solutions. Schemes based on these architectures differ in the distribution of decision making among entities involved in the control and data that is made available to these entities. Numerical analysis shows the trade-offs of proposed solutions from a performance, scalability and complexity perspectives.

KEY-WORDS: Demand Response, Direct Load Control, Smart Grids, Internet of Things, Operations Research