Optimized management of an active distribution network using AMAS combined with the RL bandit method

Sharyal Zafar

COLLEGE MATHS, TELECOMS
DOCTORAL INFORMATIQUE, SIGNAL
BRETAGNE SYSTEMES, ELECTRONIQUE

école normale supérieure rennes

THÈSE DE DOCTORAT DE

L'ÉCOLE NORMALE SUPÉRIEURE DE RENNES

ÉCOLE DOCTORALE N° 601
Mathématiques, Télécommunications, Informatique, Signal, Systèmes, Électronique
Spécialité : *Génie Électrique*

Par
Sharyal Zafar

Optimized management of an active distribution network using AMAS combined with the RL bandit method

Thèse présentée et soutenue à École Normale Supérieure de Rennes, le 15 Décembre 2023
Unité de recherche : Laboratoire SATIE

Rapporteurs avant soutenance :

M. Bruno FRANCOIS Professeur, École Centrale de Lille
M. Gauthier PICARD Directeur de Recherche, ONERA

Composition du Jury :

Président :	M. Marc PETIT	Professeur, CentraleSupélec
Rapporteurs :	M. Bruno FRANCOIS	Professeur, École Centrale de Lille
	M. Gauthier PICARD	Directeur de Recherche, ONERA
Examinateurs :	M. Marc PETIT	Professeur, CentraleSupélec
Dir. de thèse :	M. Hamid BEN AHMED	Maître de Conférences HDR, ENS Rennes
Co-enc. de thèse :	Mme Anne BLAVETTE	Chargé de recherche CNRS (SATIE), ENS Rennes
Co-enc. de thèse :	M. Guy CAMILLERI	Maître de Conférences, Université de Toulouse

Invité(s) :

M. Raphaël FÉRAUD Ingénieur de recherche, Orange Labs

Acknowledgements

First and foremost, I extend my deepest appreciation to my parents for their unwavering support, encouragement, and the countless sacrifices they made throughout my educational journey. Your support and understanding have been my constant motivation.

I am profoundly indebted to my thesis director, Hamid Ben Ahmed, for his exceptional guidance, unending patience, and genuine kindness throughout the entire thesis process. Your mentorship has been invaluable, and I am grateful for your trust in me. I also want to acknowledge my co-supervisors, Anne Blavette and Guy Camilleri, for their continuous support, availability, and for fostering an environment where I felt encouraged to explore my ideas. Your insights and guidance have been instrumental in shaping this thesis.

A special thanks goes to Raphaël Féraud for his generosity in sharing his time and expertise. Your hospitality at Orange Labs in Lannion and introduction to the world of Bandits have significantly enriched my research experience.

I extend my appreciation to the rapporteurs and jury members who took a keen interest in my work and generously contributed their expertise. Your feedback and corrections have immensely improved the quality of my thesis. I am grateful to ENS Rennes and the Regional Council of Brittany for providing the funding that supported my research. Additionally, I would like to thank the dedicated staff at ENS Rennes for their role in ensuring the smooth operation of this thesis.

To my colleagues in the Mechatronics Department, it has been a pleasure to exchange ideas, collaborate on common subjects, and share coffee breaks with you. Your camaraderie has made this journey even more rewarding. Lastly, I want to acknowledge all those whose names may not appear here but have played a role, big or small, in my academic and personal growth. Your support has been invaluable. Thank you all for being part of this incredible journey, and for contributing to the realization of this thesis. Your support and belief in me have made all the difference.

Contents

Nomenclature

Acronyms

AI Artificial intelligence

AMAS Adaptive multi-agent system

BP Balance perimeter

BRP Balance responsible party

CMAB Combinatorial multi-armed bandit

DER Distributed energy resource

DSO Distribution system operator

EV Electric vehicle

EXP3 Exponential-weight algorithm for exploration and exploitation

G2V Grid-to-vehicle

LVTF Low voltage test feeder

MAS Multi-agent system

MILP Mixed integer linear programming

ML Machine learning

NCS Non-cooperative situation

PV Photovoltaic

QCP Quadratic constrained programming

RES Renewable energy source

SoC State of charge

SoH State of health

TS Thompson sampling

TSO Transmission system operator

UCB Upper confidence bound

V2G Vehicle-to-grid

Power System

$\eta_{e,a}$ Charging/discharging efficiency of electric vehicle e at bus a

$\tilde{P}(N)$ Scheduled BRP production/consumption during its N-th settlement period

$\tilde{P}_{e,a}(t)$ Forecasted electric vehicle active power at bus a

$\tilde{P}_{l,a}(t)$ Forecasted household active power at bus a

$\tilde{P}_{p,a}(t)$ Forecasted photovoltaic active power at bus a

$c(t)$ Instantaneous grid electricity cost

$C_e(t)$ Daily total charging cost of electric vehicle e

$C_{e,pu}(t)$ Daily per-unit charging cost of electric vehicle e

$E_{e,a,bat}$ Battery capacity of electric vehicle e at bus a

$E_{e,a,tp}$ Battery throughput of electric vehicle e at bus a

$e_{e,a}(t)$ Instantaneous forecast error in the forecasted electric vehicle profile at bus a

$e_{l,a}(t)$ Instantaneous forecast error in the forecasted household profile at bus a

$e_{p,a}(t)$ Instantaneous forecast error in the forecasted photovoltaic profile at bus a

$I_{ab,max}$ Rated electrical current value flowing from bus a to bus b

$I_{ab}(t)$ Instantaneous RMS electrical current flowing from bus a to bus b

$P_a(t)$ Instantaneous active power at bus a

$P_{a,dem}(t)$ Instantaneous active power demanded at bus a

$P_{a,gen}(t)$ Instantaneous active power generated at bus a

$P_{ab}(t)$ Instantaneous active power flowing from bus a to bus b

$P_{BRP}(t)$ Instantaneous BRP production/consumption in BRP's perimeter

$P_{e,a,max}$ Maximum active charging/discharging power of electric vehicle e at bus a

$P_{e,a,min}$ Minimum active charging/discharging power of electric vehicle e at bus a

$P_{e,a}(t)$ Instantaneous electric vehicle active power at bus a

$P_{l,a}(t)$ Instantaneous household active power at bus a

$P_{p,a}(t)$ Instantaneous photovoltaic active power at bus a

$Q_a(t)$ Instantaneous reactive power at bus a

$Q_{a,dem}(t)$ Instantaneous reactive power demanded at bus a

$Q_{a,gen}(t)$ Instantaneous reactive power at bus a

$Q_{ab}(t)$ Instantaneous reactive power flowing from bus a to bus b

$SoC_{e,a,depart}$ Required minimum state of charge of electric vehicle e at bus a

$SoC_{e,a,max}$ Rated maximum state of charge of electric vehicle e at bus a

$SoC_{e,a,min}$ Rated minimum state of charge of electric vehicle e at bus a

$SoC_{e,a}(t)$ Instantaneous state of charge of electric vehicle e at bus a

$SoC_{e,a}(t_{e,a,depart})$ State of charge of electric vehicle e at bus a at its departure time

$SoH_{e,a}(t)$ Instantaneous state of health of electric vehicle e at bus a

$V_a(t)$ Instantaneous RMS voltage at bus a

$V_{a,max}$ Upper voltage limit at bus a

$V_{a,min}$ Lower voltage limit at bus a

Y_{ab} Admittance of electrical line connecting bus a and bus b

Adaptive multi-agent system

Cr_{ant} Received instantaneous antagonist criticality

$Cr_{b,a}(t)$ Bus agent's criticality (for electrical bus a)

$Cr_{brp,a}(t)$ BRP agent's criticality

$Cr_{e,a}(t)$ Electric vehicle agent's criticality (for electrical vehicle at bus a)

$Cr_{l,ab}(t)$ Line agent's criticality (for electrical line between bus a and bus b)

$Cr_{m,b,a}(t)$ Memory-based line criticality (for electrical bus a)

$Cr_{m,l,ab}(t)$ Memory-based line criticality (for electrical line between bus a and bus b)

$Cr_{max}(t)$ Received instantaneous maximum criticality

k_b Tuning parameter of the memory-based bus criticality

k_e Tuning parameter of the electric vehicle criticality criticality

k_l Tuning parameter of the memory-based line criticality

k_{brp} Tuning parameter of the BRP criticality

Multi-armed bandit

$[\mathcal{D}]$ Set of unknown distributions corresponding to each base arm

$[\mu]$ Set of unknown expectations corresponding to each base arm

$[m]$ Set of available m base arms

$[X(t)]$ Set of random variables corresponding to each base arm

$\mathcal{A}$ Action space of a learning agent

φ Vector of estimated instantaneous photovoltaic production

θ Vector of unknown learning parameters

a Selected action

a^* Optimal action

$congl(t)$ Instantaneous binary congestion flag of line agent l

$N_t(a)$ Number of times action a has been played till round t

$P(T)$ Precision of a bandit algorithm after T rounds

$Q_t(a)$ Expected reward of playing action a in round t

$R(T)$ Regret of a bandit algorithm after T rounds

$r(t)$ Observed reward in round t

$r_\mu(S(t))$ Observed reward of playing super arm $S(t)$

$rew_{e,b}(t)$ Instantaneous reward for electric vehicle e generated by line agent b

$rew_{e,i}(S_e(d))$ Reward for electric vehicle e for playing base arm i on day d

$rew_{e,l}(t)$ Instantaneous reward for electric vehicle e generated by line agent l

$S(t)$ Played super arm in round t

S^* Optima super arm

$X_i(t)$ Random outcome of i-th base arm in round t

List of Figures

List of Tables

List of Algorithms

Introduction

Context and motivation

In recent years, a strong push towards increasing the share of renewables in the total energy mix has been observed [89], [65]. This quest towards increased renewables-based energy generation is an effect of a number of causes, such as environmental concerns and economic benefits [10]. The most prominent driving factor toward sustainable smart grids has been the expected environmental benefits [78]. Increased atmospheric carbon levels can result in extreme heatwaves, floods, and air pollution which can lead to agricultural losses and biodiversity losses [43]. Economic impacts can also be expected due to increased carbon levels [174]. It has been estimated by the International Monetary Fund (IMF) agency that the impact of climate change on the world's economy is expected to reach $1.5 trillion by 2030 [91]. Thus, the EU has put forward the goal of achieving climate neutrality (net zero greenhouse gas emissions) by 2050. The EU has also set an intermediate goal of reducing its emissions by at least 55% by 2030, compared to 1990 levels [60].

The energy sector is the leading contributor toward global greenhouse gas emissions. In 2019, the energy sector was responsible for approximately 73% of global greenhouse gas emissions estimated by the IEA [86]. Fossil fuels (coal, oil, and natural gas) were responsible for about 80% of total energy-related emissions [87]. Renewable energy sources (RES) are recognized as a pivotal means to attain the targeted reduction in greenhouse gas emissions. These energy sources are also termed as sustainable energy sources. A sustainable energy source can be defined as a source that would produce energy without any harmful direct emissions, such as solar energy, wind energy, hydro-power etc. The European Union (EU) has set a specific objective to increase the proportion of renewable energy in its energy portfolio to a minimum of 32% by the year 2030 [64]. Consequently, there has been a significant rise in the integration of renewable energy sources with the existing power grid. This increased integration has also been facilitated by significant cost reductions in the manufacturing of renewable sources and storage. According to the IEA, the cost of solar panels was reduced by almost 82% in the last decade which resulted in an annual solar power capacity increase of 35% during that period [89]. This trend of increasing solar panels is expected to continue in the future as well. The total solar capacity is expected to grow from 600 GW in 2020 to 3,000 GW in 2030 as predicted by the International Renewable Energy Agency (IRENA) [93].

The progressive electrification of the transportation sector is also facilitating the replacement of conventional fossil fuel-powered vehicles with electric vehicles, thus contributing to zero direct emissions. A significant decrease in the cost of lithium-ion batteries used in electric vehicles has been helping this transition. This cost has decreased by almost 87.45% in the last decade, according to Bloomberg [27]. Environmental concerns combined with this reduction in the battery cost were among the major driving factors that increased the number of electric vehicles from 17,000 in 2010 to 10 million in 2020, according to the IEA [88]. Furthermore, these numbers

are only expected to increase. The total number of electric vehicles on the road is expected to be 145 million in 2030 by the IEA [88].

These above-discussed new grid elements are surely expected to increase its sustainability. However, they can have several undesirable impacts on the existing electrical infrastructure as well due to their uncertain and distributed nature. The existing electrical infrastructure was initially designed for uni-directional power flows (i.e., from power plants to consumers). However, the integration of a large-amount of photovoltaic (PV) panels and electric vehicles (EVs) connected to the existing electrical networks result in bi-directional power flows. Furthermore, these new grid-connected elements are uncertain and intermittent in nature [73]. Thus, they may result in unexpected voltage fluctuations beyond an allowed manageable range, thus impacting the quality of electricity supply to consumers [146]. Similarly, if a large number of EVs are charging simultaneously, it can create a significant demand for electricity. This demand could cause congestion in electrical lines, especially if the demand coincides with peak hours [99]. These congestions can also interrupt the supply of electricity to the consumers as well as can cause rapid degradation of electrical network's infrastructure [74]. Thus, the evolution of existing electrical networks becomes a necessity to tackle the mentioned challenges. Grid reinforcement (i.e., upgrading grid infrastructure) can be one of the solutions. However, downsides of grid reinforcement solutions include high cost and long lead times [80].

On the other hand, smart grid solutions have sparked a good amount of interest in the past two decades. Smart grid solutions involve the utilization of existing flexible grid elements combined with advancements in digital and communication technologies to adjust the grid operation in a context of increasing uncertainty. These smart grid solutions can be considered efficient alternatives to earlier-discussed grid reinforcement solutions [53], [18], [111]. That is why researchers from all around the world have been investigating different ways to optimally manage smart grids. Furthermore, smart grid technologies have received favorability both in monetary terms (more than $4 billion in funding for smart grid projects through the American Recovery and Reinvestment Act (ARRA) [9]) as well as in regulatory terms (Smart Grid Policy Statement [177], Order 1000 [178], and European Commission's Smart Grid Mandate [59]). The topic of this dissertation falls also in this category. In particular, it deals with the design of an optimized real-time control system applicable to a practical (large-scale) smart grid operated under uncertainty.

Work and contributions

This thesis aims to design a decentralized optimal energy management system to manage a large-scale network in real-time under uncertainties. Initially, Chapter 1 presents a detailed literature review of existing smart grid control algorithms. The first main technical part of the thesis, detailed in Chapter 2, focuses on the design of an adaptive multi-agent system for real-time grid balancing in smart grids. This system is composed of reactive agents that react to instantaneous changes in their environment to optimize energy flows in a smart grid. These agents undergo cooperative interactions with their neighboring agents to achieve the desired goal(s) of the system. The performance of this system is evaluated through simulation-based experiments in Chapter

3, by comparing it with other baseline control strategies. These baseline strategies include uncontrolled strategy and mixed-integer linear programming optimization strategy. The comparison is made in terms of optimality as well as in terms of computational requirements. The impact of smart grid size and uncertainties on the system's performance is also investigated. A detailed evaluation shows that the modeled system succeeds in controlling a large-scale smart grid in a near-optimal way.

However, the system's performance can still be improved by integrating anticipation capabilities into the agents. The second technical part of the thesis, presented in Chapters 4 and 5, implements learning capabilities to help agents make better decisions under uncertainty in their environment. Reinforcement learning algorithms, in particular multi-armed bandit algorithms, are used to enable agents to acquire learning functions. The main contribution of these chapters is the use of the multi-armed bandit theory in conjunction with an adaptive multi-agent system framework to develop a decentralized system that optimises a large-scale smart grid under uncertainty. Optimization performance is assessed by comparison with benchmark algorithms for optimizing the smart charging of electric vehicles. The comparison shows that the developed system is capable of achieving the desired objective of optimal, decentralized real-time management of a large-scale smart grid in the presence of uncertainties. Afterwards, the conclusion of this thesis and possible perspectives are discussed.

The main contributions of this thesis are as follows:

- A decentralized system using the framework of adaptive multi-agent systems is presented to optimize energy management in smart grids. The energy management system designed is fully decentralized, scalable, real-time, near-optimal and model-free (i.e., it does not require a specific distribution network model to operate). The problem of providing ancillary services to balance responsible parties by controlling flexible electric vehicles is studied to evaluate the performance of this developed system. The performance of this system is also tested in the case of pseudo-stochasticity of solar photovoltaic energy production. Work regarding this contribution is presented in Chapters 2 and 3.

- The framework of combinatorial multi-armed bandit learning is combined with the philosophy of adaptive multi-agent systems to propose a decentralized energy management system in large-scale smart grids that would reduce the impact of stochasticity on system optimality. The problem of large-scale intelligent charging of electric vehicles (>10,000 electric vehicle agents) in the presence of uncertainties is studied in order to evaluate the performance of the proposed system. Uncertainties linked to photovoltaic energy production and to agents' actions in a multi-agent environment are taken into account in this study. The impact of choosing different multi-armed bandit learning strategies on system performance is also presented in this thesis. The decentralized system based on reinforcement learning presented is near-optimal, scalable, can operate in real time, can cope with real-life uncertainties, is fair and does not require an environment model. It can be applied to control a variety of network elements (e.g. electric vehicles, electric heating/cooling equipment, distributed energy resources, etc.) at different levels (residential distribution, commercial distribution, transmission, etc.). In addition, thanks to the potential faster convergence

of multi-arm bandit algorithms compared to other commonly used algorithms such as DQN learning, the proposed system can also bring significant techno-economic benefits in online smart grid applications. The work pertaining to this contribution is elaborated in Chapters 4 and 5.

To summarize a smart grid control system combining the theories of adaptive multi-agent systems and multi-armed bandit is presented in this thesis. The final proposed system is fully decentralized, real-time, scalable, near-optimal, and adaptable. It operates without the need for a central decision-making entity, allowing each agent in the system to make its own decisions in real-time. This system is designed to handle large-scale smart grids, providing near-optimal solutions even in the presence of stochastic conditions and various uncertainties. It takes into account different types of stochasticities that may exist in practical smart grid control, ensuring fairness among decision-making agents and satisfying both global and local constraints. Importantly, it does not rely on a specific model of the electrical grid, making it versatile and adaptable for controlling various flexible grid components in a smart grid environment. A number of novel research avenues can be explored as a result of the decentralized system proposed in this thesis.

Introduction en français

Contexte et motivation

Ces dernières décennies, on a observé un fort engagement en faveur de l'augmentation de la part des énergies renouvelables dans le mix énergétique [89], [65]. Cette quête d'une production d'énergie plus renouvelable est le résultat de plusieurs facteurs, tels que les préoccupations environnementales et un coût de production devenant compétitif par rapport aux sources d'énergie non-renouvelables [10]. Les préoccupations environnementales ont représenté l'élément moteur qui a déclenché l'établissement d'objectifs environnementaux quantitatifs, entrainant dans leur sillage une réduction des coûts via des efforts collectifs au niveau du monde académique, industriel, etc. [78]. Introduire plus d'énergies renouvelables dans le mix énergétique vise à réduire les émissions de gaz à effet de serre causant un changement climatique sur notre planète. L'augmentation des niveaux de gaz à effet de serre dans l'atmosphère entraîne en effet divers phénomènes, notamment des vagues de chaleur extrêmes, des inondations ce qui entraîne des pertes humaines, des pertes agricoles et des pertes de biodiversité [43]. Des impacts économiques sont également attendus en raison de l'augmentation des niveaux de carbone [174]. Le Fonds monétaire international (FMI) estime en effet que l'impact du changement climatique sur l'économie mondiale devrait atteindre 1,5 milliards de dollars d'ici 2030 [91]. Afin de lutter contre le changement climatique, l'Union Européenne (UE) s'est fixée l'objectif d'atteindre la neutralité climatique (émission nette de gaz à effet de serre nulle) d'ici 2050. L'UE s'est également fixée un objectif intermédiaire de réduire ses émissions d'au moins 55% d'ici 2030 par rapport aux niveaux de 1990 [60].

Le secteur de l'énergie est le principal contributeur aux émissions mondiales de gaz à effet de serre. En 2019, le secteur de l'énergie était responsable d'environ 73% des émissions mondiales de gaz à effet de serre, selon l'Agence Internationale de l'Energie (AIE) [86]. Les combustibles fossiles (charbon, pétrole et gaz naturel) étaient responsables d'environ 80% des émissions totales liées à l'énergie [87]. Les sources d'énergie renouvelable (SER) sont reconnues comme un moyen essentiel d'atteindre la réduction ciblée des émissions de gaz à effet de serre. On les appelle également sources d'énergie durables. Une source d'énergie durable peut être définie comme une source qui produit de l'énergie sans émissions directes nuisibles, telles que l'énergie solaire, l'énergie éolienne, l'hydroélectricité, etc. L'Union Européenne (UE) s'est fixé un objectif spécifique d'augmenter la proportion d'énergie renouvelable dans son portefeuille énergétique à un minimum de 32% d'ici 2030 [64]. Cela a eu pour effet d'augmenter significativement la part des sources d'énergie renouvelable dans le mix énergétique. Cette intégration accrue a également été facilitée par des réductions significatives des coûts de fabrication des sources d'énergie renouvelable et du stockage. Selon l'AIE, le coût des panneaux solaires a été réduit de près de 82% au cours de la dernière décennie, ce qui a entraîné une augmentation annuelle de la capacité de production d'énergie solaire de 35% au cours de cette période [89]. Cette tendance à la hausse de l'installation de panneaux solaires devrait se poursuivre à l'avenir. La capacité solaire totale devrait

passer de 600 GW en 2020 à 3 000 GW en 2030, selon les prévisions de l'Agence Internationale pour les Energies Renouvelables (IRENA) [93].

L'électrification progressive du secteur des transports facilite également le remplacement des véhicules conventionnels à moteur à combustion par des véhicules électriques, contribuant ainsi à la suppression d'émissions directes de gaz à effet de serre. Une diminution significative du coût des batteries lithium-ion utilisées dans les véhicules électriques contribue fortement à cette transition. Ce coût a diminué de près de 87,45% au cours de la dernière décennie, selon Bloomberg [27]. Les préoccupations environnementales combinées à cette réduction du coût des batteries ont été parmi les principaux facteurs qui ont fait passer le nombre de véhicules électriques de 17 000 en 2010 à 10 millions en 2020, selon l'AIE [88]. De plus, on s'attend à ce que ces chiffres continuent d'augmenter à l'avenir. Le nombre total de véhicules électriques sur la route devrait atteindre 145 millions en 2030 selon l'AIE [88].

Ces nouveaux éléments de réseau (sources d'énergie renouvelables, véhicules électriques) devraient certainement augmenter sa durabilité. Cependant, ils peuvent également avoir plusieurs impacts indésirables sur l'infrastructure électrique existante en raison de leur nature incertaine et distribuée. L'infrastructure électrique existante a été initialement conçue pour des flux de puissance unidirectionnels (c'est-à-dire des centrales électriques vers les consommateurs). Cependant, l'intégration d'un grand nombre de panneaux photovoltaïques (PV) et de véhicules électriques (VE) connectés aux réseaux électriques existants entraîne des flux de puissance bidirectionnels. De plus, ces nouveaux éléments connectés au réseau sont incertains et intermittents par nature [73]. Ainsi, ils peuvent entraîner des fluctuations de tension inattendues dépassant une plage autorisée, ce qui impacte la qualité de la fourniture d'électricité aux consommateurs [146]. De même, des congestions peuvent être créés, par exemple si un grand nombre de véhicules électriques sont en charge simultanément, cela peut créer une demande significative d'électricité, en particulier si la demande coïncide avec les heures de pointe [99]. Ces congestions pourraient également interrompre la fourniture d'électricité aux consommateurs et entraîner une dégradation rapide de l'infrastructure du réseau électrique [74]. Ainsi, l'évolution des réseaux électriques existants devient une nécessité pour faire face aux défis mentionnés. Le renforcement du réseau (c'est-à-dire la mise à niveau de l'infrastructure du réseau) peut être l'une des solutions. Cependant, cette solution requiert un coût élevé et des délais de réalisation importants [80].

A l'inverse, les solutions de réseaux intelligents ont suscité un grand intérêt au cours des deux dernières décennies. Les solutions de réseaux intelligents consistent à utiliser des éléments de réseau flexibles existants combinés aux avancées en matière de technologies digitales et de communication pour ajuster l'exploitation du réseau dans un contexte d'incertitude croissante. Ces solutions de réseaux intelligents peuvent être considérées comme des alternatives efficaces aux solutions de renforcement du réseau précédemment mentionnées [53], [18], and [111]. C'est pourquoi des chercheurs du monde entier ont étudié différentes façons de gérer de manière optimale les réseaux intelligents. De plus, les technologies des réseaux intelligents ont bénéficié de faveurs à la fois en termes monétaires (plus de 4 milliards de dollars de financement pour des projets de réseaux intelligents dans le cadre de la loi américaine de relance et de réinvestissement [9]) et en termes réglementaires (déclaration de politique sur les réseaux

intelligents [177], Ordre 1000 [178], et Mandat sur les réseaux intelligents de la Commission Européenne [59]). Le sujet de cette thèse s'intéresse également aux réseaux intelligents. En particulier, il traite de la conception d'un système de contrôle au temps-réel optimisé applicable à un réseau intelligent de grande taille fonctionnant dans un contexte d'incertitude.

Travaux et contributions

Cette thèse vise à concevoir un système de gestion énergétique optimal décentralisé pour gérer au temps-réel un réseau de grande taille en présence d'incertitudes. Dans un premier temps, le chapitre 1 présente une revue détaillée de la littérature sur les algorithmes existants de contrôle des réseaux intelligents. La première partie technique principale de la thèse, détaillée au chapitre 2, se concentre sur la conception d'un système multi-agent adaptatif pour l'équilibrage au temps réel dans les réseaux intelligents. Ce système est composé d'agents réactifs qui réagissent aux changements instantanés de leur environnement pour optimiser les flux d'énergie dans un réseau intelligent. Ces agents interagissent de manière coopérative avec leurs agents voisins pour atteindre le(s) objectif(s) souhaité(s) du système. Les performances de ce système sont évaluées par le biais de simulations numériques au chapitre 3, en les comparant avec d'autres stratégies de contrôle de référence. Ces stratégies de référence comprennent une stratégie non contrôlée et une stratégie optimisée déterministe basée sur de la programmation linéaire en nombres mixtes (MILP). La comparaison est effectuée en termes d'optimalité ainsi qu'en termes d'effort calculatoire (temps de calcul et mémoire demandés). L'impact de la taille du réseau intelligent et des incertitudes sur les performances du système est également étudié. Une évaluation détaillée montre que le système développé parvient à contrôler un réseau intelligent à grande échelle de manière quasi-optimale.

Cependant, les performances du système peuvent encore être améliorées en intégrant des capacités d'anticipation dans les agents. La deuxième partie technique de la thèse, présentée aux chapitres 4 et 5, met en œuvre des capacités d'apprentissage pour aider les agents à prendre de meilleures décisions en présence d'incertitudes dans leur environnement. Des algorithmes d'apprentissage par renforcement, en particulier des algorithmes de bandits manchots, sont utilisés pour permettre aux agents d'acquérir des fonctions d'apprentissage. La principale contribution de ces chapitres réside dans l'utilisation de la théorie du bandits manchots conjointement avec le cadre des systèmes multi-agent adaptatifs pour développer un système décentralisé qui optimise un réseau intelligent à grande échelle en présence d'incertitudes. Les performances de l'optimisation sont évaluées par comparaison avec des algorithmes de référence pour l'optimisation de la recharge intelligente des véhicules électriques. La comparaison montre que le système développé est capable d'atteindre l'objectif souhaité de gestion au temps-réel quasi-optimale et décentralisée d'un réseau intelligent à grande échelle en présence d'incertitudes. Ensuite, la conclusion de cette thèse et les perspectives possibles sont discutées.

Les principales contributions de cette thèse sont les suivantes :

- Développement d'un système décentralisé utilisant le cadre des systèmes multi-agents adaptatifs combiné à un algorithme métaheuristique réactif. Le système

de gestion de l'énergie conçu est entièrement décentralisé, évolutif, destiné à la conduite du réseau au temps-réel, quasi-optimal et sans modèle (il ne nécessite pas de modèle de réseau de distribution spécifique pour fonctionner). Le problème de la fourniture de services systèmes pour équilibrer le réseau électrique est étudié afin d'évaluer la performance du système développé. La performance de ce système est également testée dans des cas pseudo-stochastiques de la production d'énergie solaire photovoltaïque. Les travaux relatifs à cette contribution sont présentés dans les chapitres 2 et 3.

- La viabilité de l'apprentissage par renforcement multi-agents pour optimiser les flux d'énergie dans les réseaux intelligents est également étudiée dans cette thèse. Combinaison du cadre de l'apprentissage combinatoire de bandits manchots avec la philosophie des systèmes multi-agents adaptatifs afin de proposer un système décentralisé de gestion de l'énergie dans les réseaux intelligents visant à réduire l'impact des incertitudes sur l'optimalité du système. Le problème de la recharge intelligente à grande échelle de véhicules électriques (plus de 10 000 agents véhicules électriques) en présence d'incertitudes est également étudié pour évaluer la performance du système proposé. Les incertitudes liées à la production d'énergie photovoltaïque et aux actions des agents dans un environnement multi-agents sont prises en compte dans cette étude. L'impact du choix de différentes stratégies d'apprentissage par des algorithmes de bandits manchots sur la performance du système est également présenté. Le système décentralisé basé sur l'apprentissage par renforcement développé est quasi-optimal, évolutif, peut fonctionner au temps-réel, peut faire face aux incertitudes de la vie réelle, est équitable et ne nécessite pas de modèle. Il peut être appliqué pour contrôler une variété d'éléments du réseau (par exemple, les véhicules électriques, les équipements de chauffage/refroidissement électriques, les ressources énergétiques distribuées, etc.) à différents niveaux (distribution résidentielle, distribution commerciale, transmission, etc.). De plus, grâce à la convergence potentielle plus rapide des algorithmes de bandits manchots par rapport à d'autres algorithmes couramment utilisés tels que le DQN, le système proposé peut également apporter d'importants avantages techno-économiques dans les applications en ligne (au temps-réel) des réseaux intelligents. Les travaux relatifs à cette contribution sont développés dans les chapitres 4 et 5.

Pour résumer, ce mémoire présente un système de contrôle de réseau intelligent combinant les théories des systèmes multi-agents adaptatifs et des bandits manchots. Le système final proposé est entièrement décentralisé, peut fonctionner au temps-réel, évolutif, quasi-optimal et adaptable. Il fonctionne sans avoir besoin d'une entité de prise de décision centralisée, permettant à chaque agent du système de prendre ses propres décisions au temps-réel. Ce système est conçu pour gérer des réseaux intelligents à grande échelle, offrant des solutions quasi-optimales même en présence de conditions d'incertitudes diverses. Il prend en compte différents types de stochasticités qui peuvent exister dans le contrôle des réseaux intelligents, garantissant l'équité entre les agents de prise de décision et satisfaisant à la fois les contraintes globales et locales. De manière cruciale, il ne dépend pas d'un modèle spécifique du réseau électrique, ce qui le rend polyvalent et adaptable pour le contrôle de divers composants flexibles dans

un environnement de réseau intelligent. Un certain nombre de nouvelles perspectives de recherche peuvent être explorées grâce au système décentralisé proposé dans ce manuscrit de thèse.

Chapter 1

State-of-the-art and scientific positioning

> To understand the things that are at our door is the best preparation for understanding those that lie beyond.
>
> Hypatia

Summary

This chapter starts by presenting the evolution of existing power systems. This evolution would result in a greater number of variable distributed energy resources in future power systems. Thus, smart grid optimal control systems will be an integral part of future smart grids to tackle the increase in variability due to a higher share of distributed energy resources in the overall energy mix. A number of existing smart grid optimal control solutions are presented in this chapter. These solutions are categorized based on their architecture, i.e., centralized, hierarchical, or decentralized. The rationale behind the selection of a decentralized architecture in this thesis is also explained in this chapter. Finally, the contribution of the proposed decentralized control solution compared to existing solutions is highlighted.

Contents

1.1 Power systems

One of the most complex man-made structures is often thought to be modern power systems. This can be attributed to the fact that managing power systems effectively is difficult due to the complex interactions between a variety of elements and the vast geographic areas they cover. Initially designed for lighting purposes, power systems have been one of the core drivers of progress in the late-modern age. Their applications span a broad category of sectors, such as transportation, industry, agriculture, communication, healthcare, residential real estate, etc. Electrical networks have proven themselves to be one of the fundamental pillars for the functioning of our modern society. However, power systems have not always looked the way they do now. In fact, like a living being, power systems have also been continuously evolving since their inception. A brief timeline of the history of power systems is presented in Figure 1.1. Its evolution so far can be roughly categorized into four different phases:

- **Early power systems:** During the first phase (−1890s), the foundation of power systems was made. Initially, the focus was on the design of basic building blocks of power systems which would lead to their practical adoption. Power systems in the early days of electricity were primarily used for lighting and ran on a *direct current* (DC) system [148]. Steam or water turbines produced electricity, which was then transmitted to consumers. These early networks were localized, with a power plant and distribution system unique to each city or town.

- **Growth of power systems:** During the second phase (1890s−1990s), expansion of the power networks based on alternating currents was observed worldwide, which resulted in one of the most complex man-made structures i.e., the modern electrical grid [148]. The *alternating current* (AC) technology made it possible to transmit electricity over long distances. The electrical grid became a complex system of generators, transmission lines, and distribution networks that enabled electricity to be transmitted over long distances and shared between regions. Significant electronics and digital advancements during this phase also led to increased efficiency of power systems through control and automation.

- **Development of DERs:** During this phase (1990s−2010s), a push towards *distributed energy sources* (DERs) was observed. These DERs included technolo-

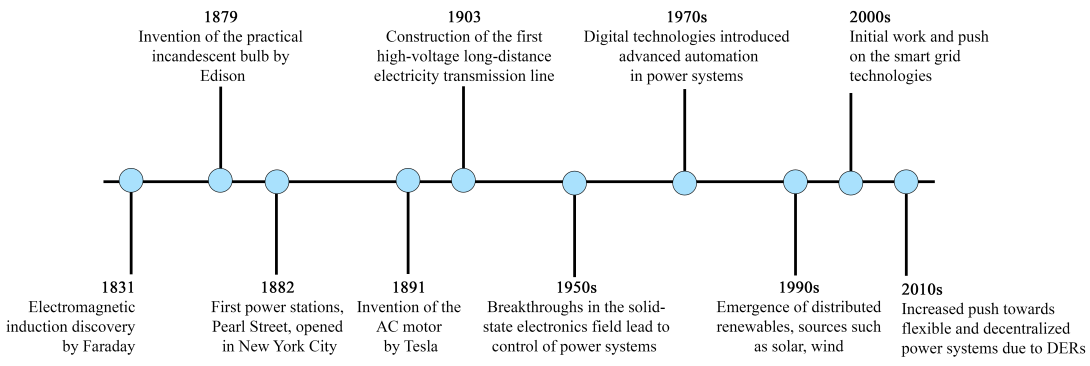

Figure 1.1: A brief timeline of power systems' evolution.

gies like solar, wind, and cogeneration [2]. The development of distributed generation, where electricity is produced at the point of use, such as rooftop solar panels, is a result of these energy sources. As a result, the conventional model of centralized power generation has been put to the test, and smart grids and microgrids, which can control the flow of electricity from various sources and loads, have been proposed.

- **Digitalization of the grid:** The most recent phase (2010s−) of the power systems' evolution is more focused on the *digitalization* of the existing electrical grid [50]. Today's advanced metering infrastructure, intelligent sensors, and other novel digital technologies have made it possible to monitor and control the electrical grid in real-time, resulting in a growing degree of digitalization of the grid. This push towards digitalization is made to increase its efficiency, reliability, resilience, and safety.

By contrasting the architectures of past power systems with those of anticipated future power systems, one can better comprehend the ongoing metamorphosis of power systems. This would not only help to better understand the historical context of contemporary power systems, but it would also highlight potential difficulties that new power systems might run into in the future and how they might overcome those difficulties by utilizing novel technologies. The following section presents this comparison.

Power systems of the past

The simplified architecture of the past power systems is shown in Figure 1.2 [85]. Historically, power systems operated with a unidirectional flow of electricity adhered to a centralized generation model. The United States Environmental Protection Agency (USEPA) defines the centralized power generation model as follows [176]:

> *"Centralized generation" refers to the large-scale generation of electricity at centralized facilities. These facilities are usually located away from end-users and connected to a network of high-voltage transmission lines.*

In the centralized generation model production of electricity was concentrated in a small number of sizable power plants. These power plants generated large amounts of electricity to satisfy the rising electricity demand. Notable examples of such centralized generation facilities include fossil-fuel-fired power plants, nuclear power plants, and hydroelectric dams. A system of high-voltage transmission lines is used to transport the power produced at these plants over great distances. For further distribution to individual consumers (industrial, commercial, or residential) via distribution lines, the voltage is stepped down at distribution substations.

In the past, power systems were typically operated by a single operator. This operator held the responsibility for ensuring the seamless and efficient functioning of the power system. The operator used *Supervisory Control and Data Acquisition* (SCADA) systems to control the power grid [15]. The utilization of SCADA systems persists in current power systems and is anticipated to remain an integral component in future

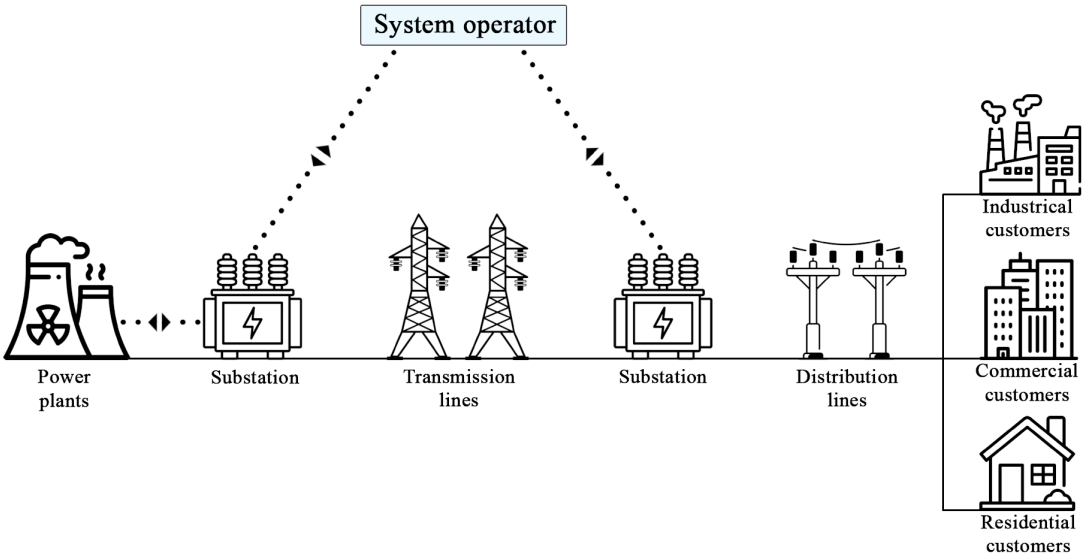

Figure 1.2: Architecture of the past power systems. Dotted lines indicate communication links and solid lines indicate electrical connections [85].

power systems. These systems are envisioned to operate in conjunction with emerging communication and control technologies, synergistically enhancing the overall functionality and efficiency of power grids. These SCADA systems are crucial in centralizing control and giving the operator real-time information. They are composed of hardware and computer software that allows for remote monitoring and management of various power grid functions. The operator can access vital data, including voltage levels, current flows, and equipment statuses, through a graphical interface, giving them a complete picture of the system. Although SCADA systems offer effective control and monitoring capabilities, the level of automation in previous power grids was primarily restricted to the transmission level [85]. In order to monitor and maintain distribution systems, which are in charge of supplying electricity to final users, there was little or no automation. Furthermore, in the centralized generation model, customers were passive recipients of electricity, lacking the ability to inject electricity back into the grid. They were billed according to their energy consumption and had limited automation capabilities for efficient control over their electricity usage.

Historically, electricity flow operated exclusively in a unidirectional manner, originating from centralized power plants and progressing toward consumers. However, this will no longer be the case in future power systems. The advent of novel grid technologies and automation will introduce prominent bidirectional power flows in future power networks. Moreover, the uncertainties linked to these novel technologies can potentially impede the intended functionality of future power systems, as elaborated in the subsequent subsections. Additionally, future power networks will be managed by a number of different operators operating in synergy, rather than being solely managed by a single entity. Consequently, future power systems will need to be smarter to effectively adapt to these substantial changes. The subsequent subsection provides a comprehensive analysis of these transformations within future power systems.

Power systems of the future: towards smarter grids

It is crucial to understand the recent and ongoing changes in power systems to envision the power grids of the future. Technical breakthroughs combined with favorable regulatory frameworks have been the key drivers for the evolution of power systems.

One of the most notable ongoing transformations in power systems pertains to a substantial increase in the proportion of renewable energy sources within the overall energy generation portfolio. This augmented share stems from the implementation of renewable technologies, including solar photovoltaic panels, wind turbines, and others, into the power systems. It is noteworthy that several of these renewable technologies had already been in existence prior to the conclusion of the 19th century (for instance, the first photovoltaic cell was invented in 1839 by a 19-year-old French physicist named Edmond Becquerel) [19]. However, it was only in the latter half of the 20th century that the collective share of renewable energy sources started its significant ascent. Two of the main catalysts behind this surge were environmental concerns and the energy crisis of the 1970s. The 1970s energy crisis had a significant impact on the adoption of renewable energy, driving governments all over the world to look for alternate and sustainable energy sources. Countries started developing laws and incentives to encourage the development and deployment of renewable energy technology as they became aware of the vulnerability of relying solely on limited fossil fuel resources [56].

The environmental concerns included both air pollution and more importantly rapid warming of the earth's temperature due to different greenhouse gas (GHG) emissions. Greenhouse gas emissions signify the release of numerous gases (such as carbon dioxide, methane, nitrous oxide, and others) through human activities, which contribute to climate change by raising the average global temperature. Since the conclusion of the Little Ice Age in the 19th century, the average global temperature has exhibited a sharp increase primarily due to the industrial revolution. Although some articles predicting the effects of global warming had emerged as early as the 20th century, genuine concerns on this subject only arose in the late 1950s when systematic measurements of rising background carbon dioxide concentrations were first initiated in 1958 [128], [162]. These environmental concerns only became stronger in the following decades, fueling research efforts in renewable energy technologies. The research and development of renewables have occupied a prominent position within the Framework Programmes (now Horizon), instigated by the European Union during the 1980s [139]. These driving factors have consequently led to a relative upswing in the techno-economic viability of many renewable energy technologies that one observes today as integral components of modern power systems.

The number of renewable energy sources connected to modern power systems is growing day by day. Nations worldwide consistently revise their greenhouse gas (GHG) emissions targets in conjunction with establishing objectives to increase the proportion of renewable energy sources within their energy portfolio. For example, the increment in the solar PV capacity of the EU since 2015 and the target set by the EU in the REPowerEU plan is shown in Figure 1.3 [62]. It can be observed that the solar PV capacity has steadily been increasing [92]. Also, the set goal according to the REPowerEU plan is to reach a solar PV capacity of 600 GW by 2030 [62]. This goal is

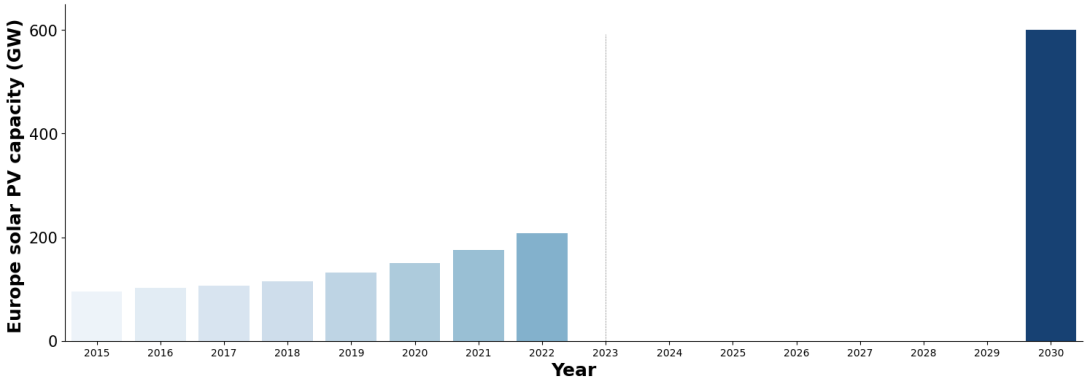

Figure 1.3: Solar PV capacity increase and the REPowerEU target [62], [92].

set by the European Union (EU) to achieve its objective of reducing its total emissions by at least 55% by 2030, compared to 1990 levels [60]. It is imperative to note that the deployment of renewable energy sources is characterized by their dispersion across the electrical grid, rather than being confined to a single location. These technologies are integrated within the grid's generation infrastructure from the high-voltage or to the low-voltage side. This enables different electricity consumers to become a more active part of the grid (i.e., prosumers). Such prosumers possess the capacity not only to buy (import) energy from the grid but also to sell (export) surplus energy back to the grid.

The electric vehicle is another technology that is becoming an increasingly integral part of the power systems as a result of the electrification of the transportation sector. Electric vehicles use electricity as their main source of propulsion, in contrast to conventional vehicles, which depend on internal combustion engines that are powered by gasoline or diesel. Although electric vehicles had early success in the late 19th and early 20th centuries, they were overshadowed by gas-powered vehicles, which offered longer trips at a lower cost due to the discovery of large petroleum sources [168]. However, electric vehicles regained attention in the late 20th and early 21st centuries due to growing environmental concerns and the energy crisis of the 1970s. Electric vehicles have several environmental advantages as they produce zero direct emissions, helping to reduce greenhouse gas emissions. They are considered eco-friendly, especially when powered by renewable energy sources. The increased focus on electric vehicles has led to significant advancements in related technologies. Additionally, the cost of lithium-ion batteries, which are crucial components of electric vehicles, has significantly decreased by around 87.45% over the past decade (as reported by Bloomberg [27]). These factors have contributed to the rapid adoption of electric vehicles since the beginning of the current century as shown in Figure 1.4. Their adoption rate is only expected to increase as electric vehicles have also been included in the REPowerEU plan by the EU. The goal is to have at least 30 million zero-emission vehicles (ZEVs) on European roads by 2030 [63].

It is also anticipated that future power systems will incorporate a variety of storage technologies to improve grid dependability, allow for the increased integration of renewable energy sources, and facilitate effective energy management [154]. Some of the most prominent energy storage technologies include lithium-ion batteries, pumped

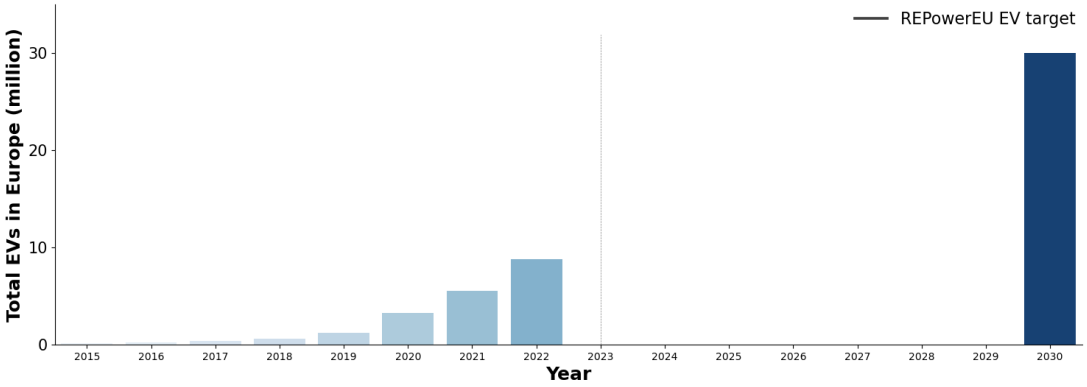

Figure 1.4: Total number of electric vehicles increase and the REPowerEU target [63].

hydro storage, hydrogen energy storage, compressed air energy storage, flywheel energy storage, thermal energy storage, and high-power super-capacitors [76]. These energy storage systems can improve grid stability and reliability by storing excess electricity during times of low demand and releasing it during times of peak demand. Energy storage can also facilitate the integration of renewable energy sources by acting as a buffer to maintain a balance between supply and demand. In future power systems, energy storage connections are expected to be spread out across the grid, much like the distribution of renewable energy sources.

The aforementioned changes relate to significant developments observed in the technological aspects of power networks. Nonetheless, these transformations extend beyond technological aspects alone, as recent regulatory and policy shifts have also had a considerable impact on power systems, influencing the reconfiguration of their control mechanisms. The liberalization of energy markets and the digitalization of power systems are leading to significant changes in the control and operation of power systems. In the past, a single utility company owned and managed all aspects of power generation, transmission, distribution, and retail services. However, with the introduction of liberalization, energy markets have been "unbundled" or separated into different entities that compete with each other, especially in wholesale and retail markets. This means that in the future, power systems will be operated by multiple operators and service providers working together to control various aspects of the grid.

The digitalization of power systems plays a crucial role in enabling efficient control of these unbundled power systems, fostering collaboration and coordination among different market participants. One key advancement in the digitalization process is the deployment of advanced metering infrastructure (AMI) including smart meters. These smart meters can provide real-time information, e.g. energy usage to both utilities and consumers, allowing for better monitoring and management of energy consumption. They also facilitate two-way communication between utilities and consumers, enabling programs like demand response and time-of-use pricing. These initiatives incentivize consumers to shift their electricity usage to off-peak hours. With the support of energy service providers, such as aggregators, customers can actively participate in programs like demand response for instance through smart and flexible devices connected to the power grid, such as electric vehicles. These advancements suggest that automation and control are expected at the distribution level (and not only the

transmission level) in future power grids.

Considering the anticipated advancements and transformations discussed earlier, one can envision a broad conceptualization of future power grids as depicted in Figure 1.5 [85]. Future power systems will undergo significant changes with the introduction of new grid elements such as renewable energy sources, electric vehicles, and energy storage systems. These elements will be dispersed throughout the power systems, transforming the way electricity is generated, consumed, and stored. Furthermore, the control and operation of future power systems will involve multiple operators working together. Aggregators managing new grid elements may communicate and collaborate with transmission and distribution system operators to ensure the reliable and efficient functioning of the power grid. This cooperative approach among operators is crucial to maintain the stability and effectiveness of the power system as it adapts to the integration of diverse technologies and distributed energy resources. The term *smart grid* is another name for this imagined power system in Figure 1.5.

Smart grid

The conceptualization of smart grids can be traced back to the latter part of the 20th century, when advancements in technology and concerns regarding the limitations of existing power systems prompted exploration into more intelligent and efficient grid solutions. However, it was in the early 2000s that the term "smart grid" gained popularity, coinciding with advancements in digital technology and communication networks that started to play a significant role in transforming power grids. Additionally, the

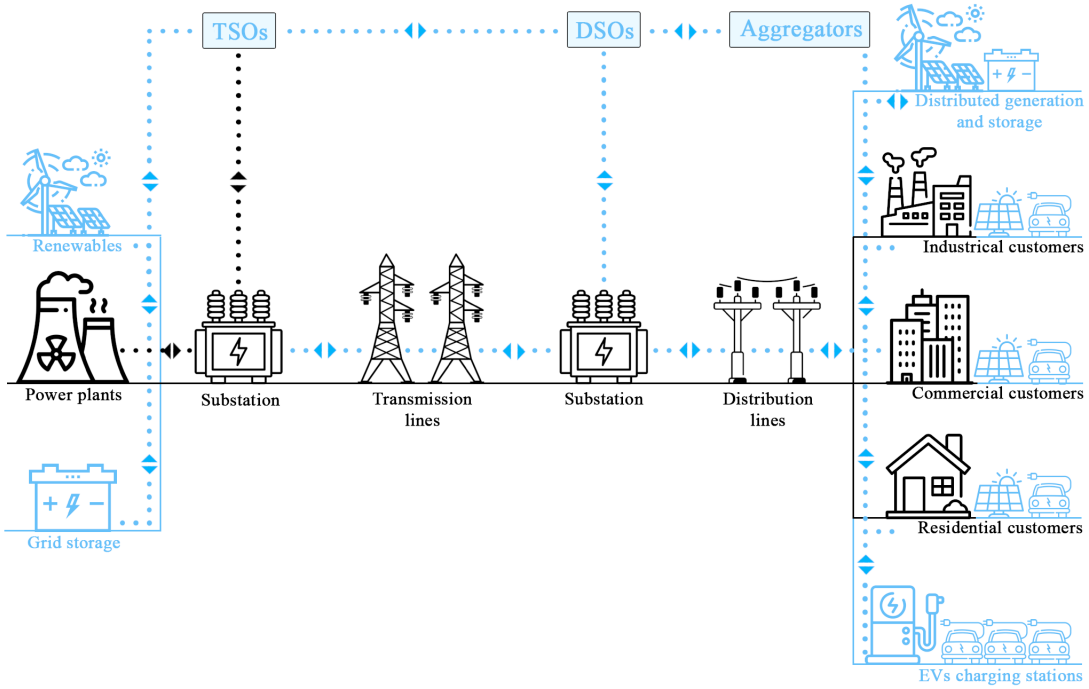

Figure 1.5: Architecture of the future power systems compared to the past power system shown in Figure 1.2. Dotted lines indicate communication links and solid lines indicate electrical connections. New grid elements are highlighted in blue.

notable blackout incident in the United States in 2003, which incurred an estimated loss ranging between $7 and $10 billion, served as a catalyst for the acceleration of smarter power system initiatives [58]. Consequently, the initial reference to "Smart Grids" within the technical literature emerged in the 2005 issue of IEEE PES Power and Energy magazine [124]. In order to gain a comprehensive understanding of the underlying principles and philosophy driving smart grids, it is crucial to establish a precise definition of the concept. There exists a number of smart grid definitions depending on the focused technologies and set objectives. However, the majority of smart grid definitions stress the significance of integrating advanced technologies and encouraging active consumer participation for a more sustainable and reliable electricity system. The International Energy Agency (IEA) defines a *smart grid* as [90]:

> *An electricity network that uses digital and other advanced technologies to monitor and manage the transport of electricity from all generation sources to meet the varying electricity demands of end users.*

On the other hand, the European Union (EU) has defined a smart grid in one of their communications as [61]:

> *An upgraded electricity network to which two-way digital communication between supplier and consumer, intelligent metering and monitoring systems have been added.*

It can be seen that the key to both of these definitions is to make use of contemporary technologies, like intelligent metering, to make it easier for consumers and system operators to communicate and automate the control. This would allow for efficient and reliable control of the power grid. There are several major factors behind the push for smart grids, including (but not limited to):

- **Environmental factors:** One of the main drivers behind the switch to smart grids has been a reduction in global emissions. Utilizing smart grids can make it possible to use environmentally friendly technologies like renewable energy sources, and electric vehicles. Lower reliance on fossil fuels and coal-fired power plants would result from increased adoption of these technologies, which would reduce emissions globally. Consequently, smart grids can aid in achieving the established carbon emission targets.

- **Increased efficiency:** Countries around the world have also established objectives to enhance their energy efficiency. For example, the European Union (EU) has set the goals of increasing electrical energy efficiency (by 32.5% by 2030, compared to 2007 levels [64]). Smart grids combined with other novel technologies such as electric vehicles are going to play a crucial role in achieving the set efficiency targets.

- **Grid stability and resilience:** Grid operators can better manage peak demand periods and maintain grid stability by adjusting consumer electricity consumption, to some extent, in a smart grid by combining advanced monitoring and control with modern technologies like demand response programs. Smart grids are expected to automatically reroute power flows in order to restore service in the event of localized outages or faults, enhancing grid resilience.

- **Economic benefits:** Smart grids are meant to facilitate the integration of renewable energy sources, such as solar and wind, which over time may lead to a reduced reliance on fossil fuels and lower energy costs. For a number of stakeholders, including energy service providers and consumers, smart grids may also present new opportunities in relation to the development of energy services and business models. Lastly, it is worth noting that smart grid technologies have the potential to mitigate the occurrence and impact of grid failures at different levels (from local failures to blackouts) and thereby reduce the associated economic losses.

Over the past two decades, there has been a lot of interest in smart grid solutions. Smart grids have received encouragement both in monetary terms (more than $4 billion in funding for smart grid projects through the American Recovery and Reinvestment Act (ARRA) [9]), and in regulatory terms (Smart Grid Policy Statement [177], Order 1000 [178], and European Commission's Smart Grid Mandate [59]). However, just like any other disruptive technology, smart grid technology has its own set of challenges as well as opportunities that can be taken advantage of to solve those problems.

Challenges and opportunities

The transition towards smart grids can pose notable challenges to the operations of electrical grids across various timescales. A brief overview of different power systems operations at different timescales along with the influence exerted by the transition to smart grids on these operations is presented in Figure 1.6 [47]. These operations encompass:

- **Capacity and operations planning:** The objective of capacity and operations planning is to make informed decisions regarding the economic viability and technological aspects related to investments aimed at enhancing and expanding generation or transmission capacity. This planning process also encompasses the strategic management of these assets over extended periods. However, infrastructure upgrades of this nature can be complex to plan and financially demanding, necessitating collaboration among multiple stakeholders. The expansion of renewable energy sources and the growing integration of distributed energy resources, such as electric vehicles and energy storage systems, has only

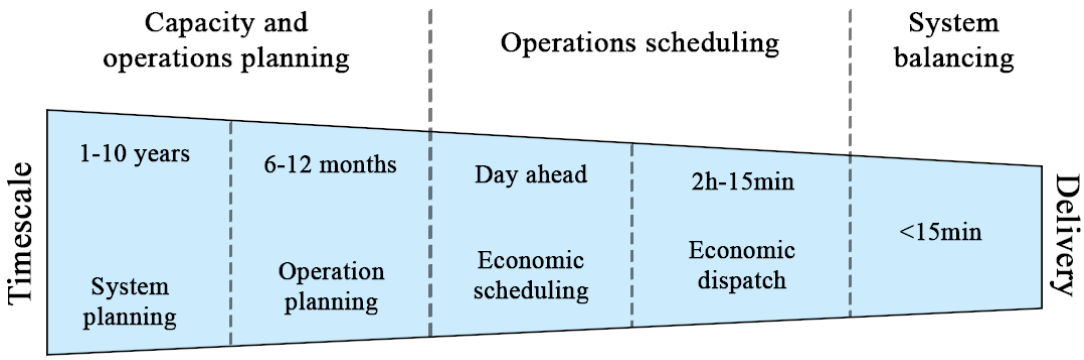

Figure 1.6: Timescales for different power system operations [47].

increased the complexity of this power system operation. It has become crucial to take these factors into account when making decisions regarding future infrastructure investments.

- **Operations scheduling:** *Day-ahead economic scheduling* and *economic dispatch* are two essential components of operations scheduling. Using demand forecasts, day-ahead scheduling determines which generators should be used for the next day up to 24 hours before delivery. Economic dispatch determines the most efficient power output of each generator that is turned on to satisfy demand using the outcomes of the scheduling process. Power system operators must incorporate forecasting information related to new renewable energy sources and distributed energy resources into their dispatching and scheduling procedures. This makes sure that the overall system operation properly takes into account and manages the variability of new grid elements. However, uncertainties associated with these new grid elements make this forecasting a challenging task. For example, the forecast for the Bonneville Power Administration (BPA) region's total solar PV production is shown in Figure 1.7 along with real-time data [29]. It can be observed that there is an error present in the forecasted day-ahead solar PV production in the BPA region compared to its actual total solar PV production.

- **System balancing:** System balancing occurs just before delivery and depends either on ancillary markets or on operational reserve. Real-time imbalances between electricity supply and demand may arise due to the uncertain nature of increasing renewable energy generation. These fluctuations in renewable energy sources can strain the grid infrastructure by causing sudden shifts in energy supply, leading to voltage fluctuations, electrical congestion, poor power quality, and a threat to the grid's stability. Additionally, the rising demand for electric vehicles can contribute to these adverse grid conditions through *peak load demands* [169]. A peak load demand occurs when a large amount of load power is drawn from the grid during certain time periods of the day.

This thesis manuscript has a specific focus on addressing challenges encountered

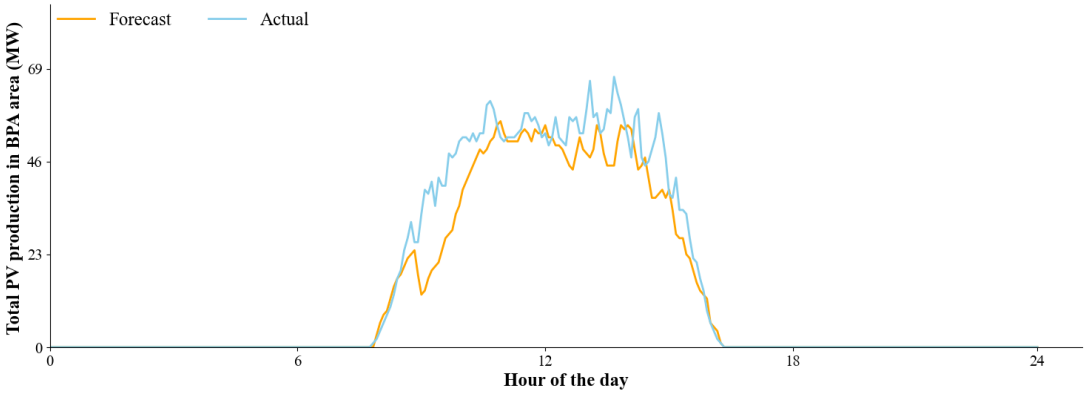

Figure 1.7: Real-time total solar PV production (on 2023-01-01) and its day-ahead forecast in the BPA area [29].

within the timescale of real-time operations in power systems, distinct from system planning or day-ahead operations scheduling. More specifically, this thesis focuses on the challenges that may arise on the distribution side as a result of the aforementioned changes in power systems. The high penetration of distributed energy resources on the distribution grid side may cause electrical current congestion, violations of voltage limits, and system imbalances. Control solutions tackling these challenges have emerged as a viable alternative to mitigate the drawbacks associated with grid reinforcement measures [80]. By integrating control solutions with energy storage systems, operators and service providers gain the ability to effectively regulate and maintain the desired operation of the electrical grid. Battery energy storage systems, in particular, offer a promising means to alleviate the impact of renewable energy variability on power system operations. These systems can store excess energy during periods of high generation and subsequently release it as needed.

Additionally, modern electric vehicles, equipped with lithium-ion batteries, are going to play a pivotal role in future power grids. Leveraging the higher discharge rates of lithium-ion batteries, which outperform other energy storage methods such as compressed air or pumped hydro with slower ramping rates (on the order of hours), electric vehicle batteries can also be utilized for grid support operations. The control of electric vehicles can aid in mitigating peak load demands, thereby contributing to grid stability. Given these inherent control capabilities, electric vehicles are often referred to as flexible entities within future power systems. It is crucial to emphasize that the smart grid control system's design must adhere to the constraints imposed by diverse market actors operating at different levels (such as distribution system operators, prosumers, etc.). Consequently, the system must not compromise the performance of one market actor while striving to attain the objectives of another. A careful balance is necessary to ensure harmonious functionality across an entire smart grid. This complexity enhances the interest and significance of optimizing control algorithms in this domain, making it a compelling research topic for scholars worldwide. The following section provides a comprehensive comparison of various smart grid control algorithms proposed in the existing literature.

1.2 Smart grid control

This section presents a compilation of literature studies centered around smart grid control. Depending on the particular operation being studied, smart grid control strategies may employ different decision timescales, such as day ahead, intra-day, and real-time. However, the primary focus of this thesis lies in the timescale related to real-time operations, as mentioned earlier. As a result, only literature studies within this specific timescale category are included in this section. Nonetheless, it is worth mentioning that the literature also encompasses numerous scientific studies that address other categories of power system operations, such as operations scheduling performed in [192], [106], [66], and [193]. Moreover, since this thesis concentrates on the control of electric vehicle battery charging to enhance grid flexibility, this section exclusively showcases smart grid control solutions that specifically address the management of electric vehicle charging or battery storage systems to facilitate grid services. However, it

should be noted that although this manuscript focuses on the design and testing of a system tailored to control electric vehicle charging, it is important to note that the system possesses adaptability. Consequently, it can be customized to manage other flexible components within future smart grids. The literature studies presented in this section can be classified into different categories, including system architecture, methodology, application under study, experimentation, and more. Given the main objective of this thesis to develop a practical smart grid control system based on decentralization (for reasons explained later in this chapter), the proposed system's architecture has been used to categorize the literature studies. These studies can be grouped into three designated categories as listed below:

- Centralized
- Hierarchical
- Decentralized

Centralized systems

A centralized system can be defined as a type of control system in which the decision-making process is done by only a single central authority. An illustration of this type of control is shown in Figure 1.8. It can be seen in Figure 1.8 that only a single entity (highlighted in blue) is responsible for making decisions on behalf of entities present in the system when the control is centralized. This central node (entity) will gather information from all nodes (entities) of the system, perform its centralized decision-making process, and then communicate the results to all other nodes in the system. Decisions may be made more efficiently (in terms of optimality) when decision-making authority is centralized compared to non-centralized methods. It removes the need for exhaustive coordination among different system entities, which may lead to control solutions with better optimality. This particular control approach has been extensively investigated and analyzed for its applicability in power system control applications.

Numerous smart grid optimal control solutions based on centralized architecture have been proposed in past years. This thesis manuscript will discuss some of them here. In [179], two rule-based real-time algorithms and one linear programming-based algorithm have been proposed to increase the self-consumption of PV generation through controlling EV charging. The correlation of transport systems and electrical networks to optimize the charging of EVs is explored in [204]. A novel binary swarm optimization algorithm to solve the unit commitment problem and competitive swarm optimization for demand side management in the presence of EVs is presented in [190]. A smart grid control system to perform system balancing utilizing electric vehicles has been proposed in [83]. In [196], another energy management system to maintain the grid balance by controlling electric vehicles' charging/discharging has been presented. Centralized control strategies to minimize the total charging costs of electric vehicles are also proposed in [194] and [127]. Charging cost minimization of EVs has also been achieved using stochastic mixed integer linear programming. In [149], a stochastic mixed integer linear programming-based algorithm has been suggested to

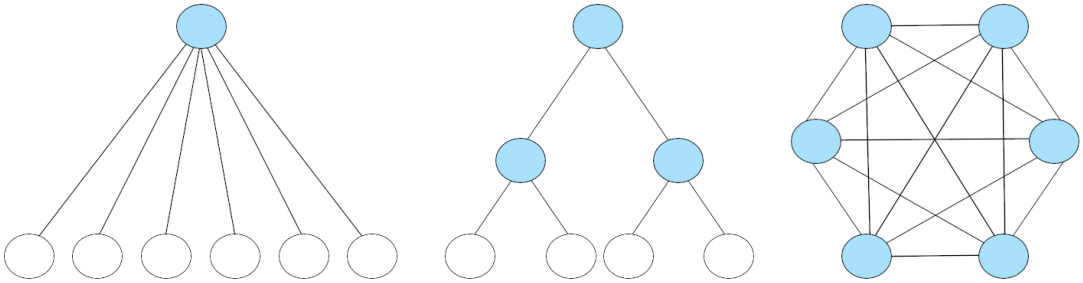

Figure 1.8: Centralized architecture (left), hierarchical architecture (middle), and the decentralized architecture (right) illustrations. Solid lines indicate the exchange of information.

optimize energy flows in smart buildings with a battery energy storage system. A control algorithm to minimize energy losses in smart grids by controlling the charging of electric vehicles has been given in [14]. The impact of real-life stochasticities on grid balancing has been studied in [197], and a control algorithm to minimize real-time mismatches using electric vehicles has been proposed. In [156], a system utilizing dynamic programming is presented to optimize the energy flows by minimizing grid imports in smart grids while considering the constraints of prosumers. A MILP-based control system is presented in [32] to minimize the carbon emissions of a microgrid as well as its total operating cost. The constraints related to the electrical batteries of prosumers are also taken into account in the mentioned system. A centralized system to optimize the charging of electrical vehicles while considering prosumers and grid constraints is discussed in [144]. This system has been evaluated through both simulation and real-life case studies. A reinforcement learning methodology using the twin delayed deep deterministic policy gradient (TD3PG) and proximal policy optimization (PPO) has been applied in [145] for robust voltage control in distribution grids under stochasticity. A control system utilizing second-order cone programming to solve the short-term problem of providing flexibility to the electrical grid has been proposed in [167]. Centralized systems based on multi-armed bandit have been proposed to control the charging of electric vehicles in [117] and [199].

Although these systems are capable of fulfilling their intended functionality and achieving optimal solutions through their centralized control architectures, there are significant drawbacks associated with centralization. In a centralized control system, the central controller must gather information from all system entities before initiating the decision-making process. This can pose challenges in practical smart grids of large-scale, as it may be impacted by communication delays. Moreover, the scalability of centralized control is limited, as the time required to obtain desired control solutions can increase dramatically with a higher number of entities in a centralized system. Additionally, a single point of failure issue can arise within a centralized system. Furthermore, centralized systems raise concerns regarding data privacy, as the centralized controller requires access to private data from all entities. Consequently, such systems may not be the best option for real-world applications, particularly for the real-time control of large-scale smart grids. These aforementioned drawbacks highlighted the importance of distributing control tasks among multiple entities, prompting a shift from centralized architectures to a hierarchical approach.

Hierarchical systems

In response to the drawbacks of centralization and the desire to maximize its benefits, researchers have also explored hierarchical smart grid control solutions. Hierarchical control systems organize decision-making tasks into multiple levels within a structured hierarchy, with varying levels and functionalities based on the specific application. Each level in the hierarchy is guided by its own objectives and constraints, contributing to the overall management of the system. Figure 1.8 also provides the illustration of a hierarchical problem-solving architecture. In hierarchical systems, decision-making is distributed across multiple levels (represented by blue nodes), unlike the centralized architecture. This hierarchical problem-solving approach holds the potential to overcome the challenges of centralization. Consequently, the advantages of hierarchical architectures have piqued the interest of researchers, leading to the proposal of smart grid control solutions based on hierarchical designs.

Numerous research works have been published on hierarchical control solutions. Some of them will be detailed here. In [81], the charging behavior of electric vehicles is managed by a hierarchical control system utilizing particle swarm optimization (PSO). The studied system optimizes the total charging cost of EVs while preventing electrical congestion. Swarm intelligence involves non-centralized agents working together to achieve a common goal. Its functionality is based on interactions between many social animals such as bees, wasps, ants, bats, birds and whales. The rules defined for social interactions can be different in these methodologies but the end goal is the same i.e. convergence to an optimum [123]. Non-cooperative game theory has also been used in [203] to integrate the best allocation of battery energy storage systems on the distribution side. Non-cooperative game theory involves designing a system with multiple entities competing against each other to achieve their respective goal(s). In [187], a hierarchical system based on graph theory has been used to minimize the charging losses of the system and to support the grid using electric vehicles. In general, graph theory involves modeling a system to study the pairwise relationship among different objects. A hierarchical control system based on a combination of fuzzy decision-making using neural networks and genetic algorithm-based optimization has been proposed in [150] to maintain the desired functioning of a smart grid. Hierarchical control systems to maintain grid balance has been studied in [125] and [186]. In [75], a hierarchical system based on heuristic control has been developed to minimize the cost of supporting a microgrid using electric vehicles. To find optimal charging strategies for electric vehicles, hierarchical systems have been proposed in [100] and [142] utilizing game theory and neural networks-based decision-making, respectively. In [126], a hierarchical multi-agent system based on quadratic optimization has been presented to incorporate demand response and coordinated charging of electric vehicles in distribution networks.

While hierarchical systems have the capability to tackle the drawbacks of centralized systems up to an extent, they may not completely address all of the drawbacks of centralization. This is because hierarchical systems can still be strongly impacted by communication delays, particularly in large-scale systems. Additionally, hierarchical systems may still be susceptible to single points of failure (e.g., white nodes in Figure 1.8 are still dependent on blue nodes for their decision-making). Data privacy

concerns may persist even in a hierarchical control system. Moreover, hierarchical systems may face challenges in scaling effectively to handle complex control problems in large-scale smart grids. Considering these potential drawbacks of hierarchical systems, there has been a growing interest in fully decentralized control systems as an alternative solution.

Decentralized systems

In contrast to centralized and hierarchical systems, a decentralized control system is a system where decision-making tasks are distributed among multiple entities of the system, rather than being leveled or centralized in a single authority. The difference between decentralized architecture against centralized and hierarchical architecture is given in Figure 1.8. It can be confirmed in Figure 1.8 that all entities (nodes) of a decentralized system are participating in the decision-making process (highlighted in blue). These entities may interact with each other to facilitate their decision-making processes. In comparison to centralized or hierarchical systems, a decentralized architecture system is more likely to have a more complex set of communication needs. Therefore, the goal is often to keep these interactions to a minimum while maintaining decentralization and system performance. On the other hand, decentralization helps in tackling the drawbacks of centralization much better than hierarchy. Such systems may not suffer from scalability issues, data privacy concerns, and a single point of failure. Due to these reasons, the design of a decentralized control system is an active area of research. Decentralization in software-based smart grid control solutions can be achieved through a number of modeling paradigms. Most prominent among these modeling paradigms are peer-to-peer (P2P), blockchain, and multi-agent systems (MASs). These modeling paradigms can be combined with different optimization techniques such as dynamic programming, stochastic dynamic programming (SDP), machine learning, alternating direction method of multipliers (ADMM), and other heuristics to optimize energy flows in a smart grid.

In a classical P2P framework, a number of interconnected nodes (also known as peers) communicate directly with each other. There is no central authority in such a system. Thus, a number of researchers have utilized the P2P paradigm to solve smart grid optimal power flows in a decentralized manner. In [189] and [135], decentralized systems based on the peer-to-peer (P2P) architecture to maintain the grid balance and minimize total energy losses are presented. These systems utilize the ADMM methodology to perform decentralized optimization. Generally, ADMM is used to solve a convex optimization problem by breaking it into smaller parts. It can also be used as a heuristic to solve non-convex optimization problems. In [18], optimal power flow has been performed using the ADMM technique in a P2P market. The feasibility of the P2P market under both exogenous network charges (provided by the system operator a priori) and endogenous network charges (updated by the system operator at each iteration) has been studied in this system. The resilience of the proposed P2P system under stochastic conditions has also been discussed. The impact of asynchronous communication in a P2P market designed to minimize the total grid operational cost using the ADMM technique has been studied in [53]. In [40], the ADMM technique has also been used to control electrical congestion in a P2P market. The algorithm reduces the

number of overloaded lines as well as the amplitude of the overload while maintaining its scalability. The philosophy of P2P has also been applied in combination with dynamic programming to design an energy trading marketplace that would maximize the benefit of each agent while respecting the constraints of the electrical network [35]. A key difference between ADMM and dynamic programming optimization techniques is that the ADMM is used for problems with separable components (separable variables or constraints) while dynamic programming is used for problems that exhibit overlapping sub-problems. A combination of both of these optimization methodologies is also possible to solve smart grid optimization problems in a P2P framework. This is explained in [111] and [112] to optimize the charging of electrical vehicles.

The blockchain is another promising modeling paradigm that helps in achieving decentralization. A blockchain is decentralized and distributed ledger technology. It consists of a growing list of blocks (records) [129]. It must be noted that there exists a close relationship between P2P and blockchain technologies as a blockchain is generally managed by a P2P computer network. Thus, similar to P2P, a blockchain can also be used to perform resource sharing in a decentralized manner. However, a key difference is in the general security standards of both approaches. In a blockchain, each block is linked to its subsequent block via complex cryptographic hashes. Thus, it provides a higher level of security and trust (which generally comes at the cost of increased complexity of using cryptographic algorithms and consensus mechanisms). This increased security is particularly useful in the presence of an adversary in the system. Blockchain technology-based systems have also been proposed in the literature to perform smart grid optimal control. A blockchain-based system combined with the ADMM technique to perform decentralized optimized control of DERs is presented in [131]. In this system, the blockchain makes the system decentralized by ensuring fair energy trading without relying on a single entity. In [116], an adaptive blockchain-based electric vehicle charge control system is proposed to minimize the overall charging cost for EV users and power fluctuations in the grid. A comparison to confirm the superior performance of this developed system with a genetic optimization algorithm has also been presented. In [198], a blockchain-empowered system to optimize energy trading is presented. This decentralized system managed to reduce the user's individual cost by up to 77% and lower the overall cost by 24%. Another smart grid control system utilizing blockchain technology to maintain voltage stability in the presence of DERs is proposed in [46]. The control performed in this system is based on the proportional-fair rule.

The multi-agent system (MAS) is a prominent modeling paradigm to achieve decentralization of the system's architecture. The entities of such a decentralized system are often referred to as *agents*, and that decentralized control system is thus referred to as a *multi-agent system*. In a multi-agent system, agents interact with each other (cooperating or competing) and with their environment to achieve a desired set of goals. A key difference between a conventional P2P system and a MAS may be that in P2P the entities (peers) may be lesser autonomous compared to entities (agents) of a MAS. Furthermore, P2P systems may have limited coordination mechanisms (relying on peer-to-peer interactions for resource sharing) in contrast to MASs (in which agents can coordinate their actions to achieve desired goals). However, the boundaries of these two modeling paradigms have started to overlap significantly in recent years.

This is because the entities (peers) in the next generation of P2P networks are desired to have general properties (i.e., autonomy, reactivity, pro-activeness, self-organization, etc.) of entities (agents) in a MAS. Thus, although traditional P2P networks were primarily created for resource sharing, these systems can be designed to address complex optimal control problems for the smart grid by incorporating the characteristics of the agents in a MAS. The development of MASs to optimize energy flows in smart grids is an active area of research. Researchers have proposed a number of MASs performing the desired optimization through various techniques such as rule-based strategies, heuristics, reinforcement learning, etc.

A MAS smart grid control system to maintain power balance has been developed in [57] by defining simplistic actions for each agent type in the system. However, the constraints of the network operator have not been considered in this study. Another multi-agent system to minimize the cost of grid imbalance while avoiding electrical congestion using electric vehicles has been proposed in [182]. The same task has been achieved along with consideration of voltage stability in [137]. Although, agents in MASs of [57], [182], and [137] are highly reactive to changes in their environments but still such systems may not be able to tackle the impact of stochasticity comprehensively due to lack anticipative abilities. Hence, a good number of multi-agent systems exploiting reinforcement learning to incorporate anticipative abilities in their agents have also been proposed. A reinforcement learning-based decentralized control system for optimal charging of electric vehicles to maintain the grid's stability has been presented in [207]. However, no voltage constraints of the distribution system operator were considered in this system. In [110], a MAS using reinforcement learning has been developed to manage a DC microgrid with battery and super-capacitor storage. The goal of the system was to minimize the energy imports from the grid and utilize the storage facilities with DERs in an efficient manner. Constraints of system operators were considered in this study but the penetration of EVs in the electrical network was not included. A similar objective has been achieved using reinforcement learning as well in [113]. In this study also power grids without electric vehicle penetration were studied. Reinforcement learning has also been combined with the theory of MAS in [160] to minimize energy mismatches in a smart grid. A minority games-based MAS has been studied in [82] to optimize a smart building with battery energy storage by minimizing the power flow from the grid during peak hours and efficiently utilizing the energy generated from DERs installed in the building. However, the minority game is a highly simplified model that may not capture all the complexities of real-world decision-making scenarios.

There exists a specific class of multi-agent systems that argues to bring strong self-organization in a decentralized system in contrast to potential weak self-organization in conventional multi-agent systems. This sub-class of MASs is known as the adaptive multi-agent system (AMAS). This theory of adaptive multi-agent systems has been exploited to design a decentralized system for grid balancing in real-time while satisfying the constraints of different electricity market stakeholders [26]. Strong self-organization in adaptive multi-agent systems is a significant advantage over conventional MAS or P2P systems with weak self-organizations. This is due to the fact that strong self-organization can allow one to model and solve large-scale and complex smart grid optimization problems in a better manner. That is why the theory of AMAS

is utilized in this thesis to develop a decentralized smart grid control algorithm. This theory of AMAS is eventually combined with reinforcement learning to perform the desired stochastic optimization. Thus, the proposed system is enabled to tackle large-scale smart grid optimization problems under stochasticity. This proposed system does not require any model of its environment (i.e., electrical distribution network) which may be essential in other ADMM-based or SDP-based decentralized control systems. Furthermore, this AMAS-based system does not demand the studied smart grid optimization problem to have any specific structure (e.g., separable variables in ADMM or overlapping sub-problems in SDP) as it achieves convergence purely based on cooperation among its agents.

Table 1.1 provides a comprehensive comparison of the discussed smart grid optimal control systems. The comparison is based on the system's *architecture*, *objective*, the inclusion of *renewable energy sources* (RESs), consideration of *electrical congestion* and *voltage congestion* (i.e., voltage limits violation), and the *total number of agents* in hierarchical and decentralized systems. The table also highlights the specifications of the smart grid control systems proposed in this thesis. Detailed analysis and discussion on the comparison between the proposed systems and existing smart grid control systems, along with their novelty, are presented in the subsequent section.

Ref.	Architecture	Objective	RESs	Electrical congestion	Voltage congestion	No. of agents
[179]	Centralized	Min. grid energy import	Yes	Yes	Yes	-
[204]	Centralized	Optimize EVs charging	No	Yes	No	-
[190]	Centralized	Demand side management	Yes	No	No	-
[83]	Centralized	Grid balancing	Yes	No	No	-
[196]	Centralized	Grid balancing	Yes	Yes	Yes	-
[194]	Centralized	Min. EVs charging costs	Yes	No	No	-
[127]	Centralized	Min. EVs charging costs	Yes	Yes	Yes	-
[149]	Centralized	Optimize energy flows	Yes	Yes	Yes	-
[14]	Centralized	Min. energy losses	No	Yes	Yes	-
[197]	Centralized	Grid balancing	No	No	No	-
[117]	Centralized	Optimize EVs charging	Yes	No	No	-
[199]	Centralized	Min. EVs charging costs	Yes	No	No	-
[156]	Centralized	Min. grid energy import	Yes	No	No	-
[32]	Centralized	Min. CO2 emissions and costs	Yes	No	No	-
[144]	Centralized	Optimize EVs charging	Yes	Yes	Yes	-
[145]	Centralized	Voltage control	Yes	Yes	No	-
[167]	Centralized	Min. operational cost	Yes	Yes	Yes	-

Ref.	Architecture	Objective	RESs	Electrical congestion	Voltage congestion	No. of agents
[81]	Hierarchical	Min. EVs charging costs	Yes	Yes	Yes	28
[203]	Hierarchical	Energy storage allocation	Yes	No	No	24
[187]	Hierarchical	Min. energy losses	No	Yes	Yes	10
[150]	Hierarchical	Min. cost and emissions	Yes	No	No	8
[125]	Hierarchical	Grid balancing	Yes	No	No	100
[186]	Hierarchical	Grid balancing	Yes	Yes	Yes	5000
[75]	Hierarchical	Min. outages cost	No	No	No	25
[100]	Hierarchical	Min. EVs charging costs	No	Yes	No	1237
[142]	Hierarchical	Min. EVs charging costs	No	Yes	Yes	34
[126]	Hierarchical	Demand side management	No	No	Yes	269
[189]	Decentralized	Grid balancing	Yes	No	No	504
[135]	Decentralized	Min. energy losses	Yes	Yes	Yes	6
[18]	Decentralized	Min. operational cost	Yes	Yes	Yes	2000
[53]	Decentralized	Min. operational cost	No	Yes	Yes	31
[40]	Decentralized	Congestion control	Yes	Yes	Yes	40
[35]	Decentralized	Min. energy trading loss	No	Yes	No	2000
[112]	Decentralized	Min. grid energy import	Yes	Yes	Yes	200

Ref.	Architecture	Objective	RESs	Electrical congestion	Voltage congestion	No. of agents
[131]	Decentralized	Min. operational cost	Yes	Yes	Yes	55
[116]	Decentralized	Min. EVs charging costs	Yes	Yes	Yes	100
[198]	Decentralized	Max. individual benefits	No	No	No	18
[46]	Decentralized	Voltage regulation	Yes	Yes	Yes	7
[57]	Decentralized	Grid balancing	Yes	No	No	504
[182]	Decentralized	Grid balancing	Yes	Yes	No	202
[137]	Decentralized	Grid balancing	Yes	Yes	Yes	8
[207]	Decentralized	Optimize EVs charging	No	Yes	No	500
[110]	Decentralized	Min. grid energy import	Yes	Yes	Yes	8
[113]	Decentralized	Min. grid energy import	Yes	No	No	4
[160]	Decentralized	Grid balancing	Yes	Yes	Yes	28
[82]	Decentralized	Min. grid energy import	Yes	No	No	24
[26]	Decentralized	Grid balancing	Yes	Yes	No	55
Ch. 2&3	Decentralized	Grid balancing	Yes	Yes	Yes	495
Ch. 4&5	Decentralized	Min. EVs charging costs	Yes	Yes	Yes	10,175

Table 1.1: Comparison of different smart grid optimal control literature studies.

1.3 Scientific positioning

The scientific positioning of this thesis will facilitate the demonstration of its novelty within the context of existing knowledge. This thesis' contributions fall under the more general category of decentralized control in smart grids. If the term "decentralized control in smart grids" is entered as an input on Google Scholar, one should expect to see a clear upward trend in the total number of search results for each year. This trend is presented in Figure 1.9. This upward trend is evidence that decentralized control of smart grids is a growing area of research. This attests to the topic of this thesis' relevance, which has been garnering steadily rising attention from the international research community. A comparative analysis among existing smart grid control systems has been presented in Table 1.1. Decentralized smart grid control systems have already been proposed in the literature to satisfy the constraints of various market actors while ensuring the desired operation of a smart grid in the presence of uncertainties associated with renewable energy sources. However, the majority of these systems consider a relatively small number of electric vehicles, typically ranging from a few hundred to a few thousand. In contrast, real-life EV fleets may involve thousands or even millions of electric vehicles. The primary novelty of this thesis lies in its exploration of **large-scale** decentralized smart grid control systems operating in **real-time** and under **uncertainties**. Thus, the underlying question being tackled in this thesis is:

- Can a real-time decentralized energy management control system be designed for large-scale active electrical distribution networks, capable of effectively managing challenges caused by uncertainties in distributed energy resources by providing flexibility services to the smart grid?

To tackle the above-mentioned research question, the following hypotheses have been made:

- The theory of adaptive multi-agent systems can serve as a valuable framework for developing an effective real-time decentralized energy management control system for large-scale active electrical distribution networks.

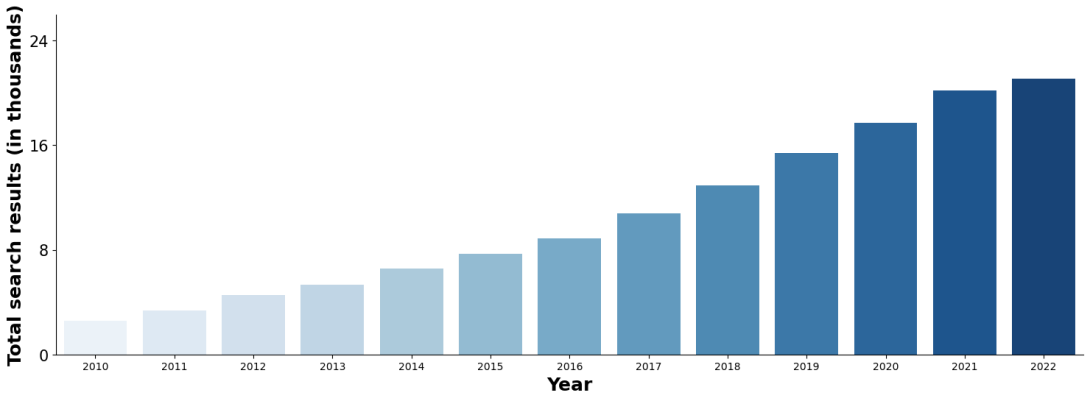

Figure 1.9: Number of search results of "smart grid decentralized control" on Google scholar.

- The incorporation of the multi-armed bandit class of reinforcement learning algorithms into the aforementioned adaptive multi-agent system can enhance the system's performance under real-life uncertainties, while simultaneously preserving its scalability and real-time operations capabilities.

The first hypothesis is grounded in the principles and practical applications of adaptive multi-agent systems. Such systems achieve the desired objective(s) through cooperative actions among agents, rather than relying solely on functional division [21]. In adaptive multi-agent systems, agents can adapt their goals based on their local environments, resulting in a more adaptable system compared to fixed-functionality multi-agent systems. Moreover, adaptive multi-agent systems typically operate in a decentralized manner with minimal time-consuming interactions limited to their local neighborhoods, ensuring the scalability and real-time capabilities of the designed system. The adaptive multi-agent system represents a particular type of multi-agent framework suitable for designing an adaptive decentralized system. While this system has the potential for scalability and real-time functionality, its overall efficiency under uncertain conditions relies heavily on the careful design and capabilities of its individual agents. Thus, the second hypothesis is made by looking at the prediction capabilities of different reinforcement learning algorithms [151]. The selection of multi-armed bandit algorithms is based on their potential advantages over standard reinforcement learning algorithms like deep Q-learning. Notably, multi-armed bandit algorithms demonstrate faster convergence, making them promising candidates for optimizing the system. Additionally, they have shown practical success in optimizing communication within the Internet of Things (IoT) devices [28], [24]. The hypothesis postulates that a well-designed multi-armed bandit algorithm can efficiently address the stochastic smart grid control optimization problem, while preserving the system's scalability and real-time operation capabilities. This thesis focuses on the validation of the made hypotheses. To do so, the following two decentralized smart grid control systems are designed and evaluated:

- A decentralized system based on adaptive multi-agent system theory to perform grid balancing in real-time while satisfying the constraints of grid operators and prosumers (Chapter 2). To evaluate the system's performance, simulation case studies with 55 and 495 electric vehicle agents have been conducted.

- A decentralized system combining adaptive multi-agent system theory with multi-armed bandit learning to perform smart charging under stochasticities while also satisfying the constraints of grid operators and prosumers (Chapter 4). To evaluate the system's performance, simulation case studies with 55 and 10,175 electric vehicle agents have been conducted.

The studied large-scale distribution network in this thesis encompasses over 10,000 electric vehicles. It should be noted that the system considers constraints from various energy market actors and addresses uncertainties related to renewable energy sources (RESs) while ensuring scalability. To achieve real-time control operations with scalability, adaptive multi-agent system concepts have been utilized. Additionally, the designed adaptive multi-agent system is combined with the combinatorial multi-armed

theory to reduce the impact of uncertainties while maintaining scalability. Combinatorial multi-armed bandit learning belongs to a simpler class of reinforcement learning, providing faster convergence compared to more complex algorithms like deep Q-learning. This faster convergence is significantly advantageous in smart grid control applications where online learning is used. As far as the author is aware, the combination of adaptive multi-agent system theory and combinatorial multi-armed bandit concepts for optimal smart grid control is a novel approach and has not been explored before.

1.4 Manuscript contributions and organization

Contributions

In this thesis, the problem of real-time energy management of large-scale future smart grids under uncertainties is studied. The major contributions of this thesis include:

- **Decentralized energy management:** A decentralized system is presented to optimize energy management in smart grids. The proposed decentralized system utilizes the framework of adaptive multi-agent systems combined with a reactive heuristic algorithm to ideally search for optimal energy management policies. The designed energy management system is fully decentralized, model-free (i.e., it does not require an accurate distribution network model for its functioning), and scalable. It can be applied to control a variety of grid elements (e.g., electric vehicles, electric heating/cooling equipment, distributed energy resources, etc.) at different levels (residential distribution, commercial distribution, transmission etc.) in real-time. However, for the sake of simplicity, only electric vehicles are considered here. This load represents the most complex ones, as it is bidirectional and has a dynamic point of connection.

- **Multi-agent reinforcement learning:** The viability of multi-agent reinforcement learning to optimize energy flows in smart grids is also studied in this thesis. The framework of combinatorial multi-armed bandit learning is combined with the philosophy of adaptive multi-agent systems to propose a decentralized smart grid energy management system that would reduce the impact of uncertainties on the optimality of the system. The impact of choosing different multi-armed bandit learning strategies on system performance is also discussed in this thesis. The presented reinforcement learning-based decentralized system is scalable, model-free, operates in real-time, and can tackle real-life uncertainties. Furthermore, combinatorial multi-armed bandit algorithms can provide faster convergence compared to more complex reinforcement learning algorithms, such as deep Q-learning. This faster convergence brings a significant economic advantage in smart grid applications where an agent is continuously learning mainly through online interactions.

The mind map of this thesis' main contributions is presented in Figure 1.10. As shown in Figure 1.10, contributions of this thesis can be divided into two notable categories along with a state-of-the-art review. The first contribution category covers

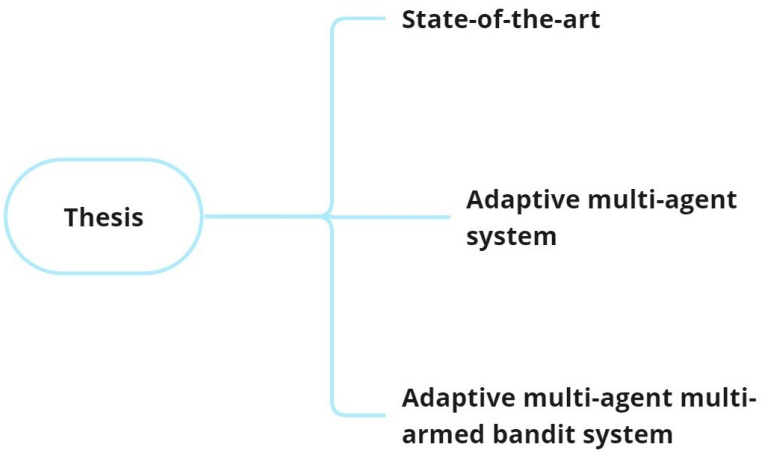

Figure 1.10: Mind map of the thesis contribution.

the design and detailed evaluation of an adaptive multi-agent system to optimize grid balancing in smart grids. The numerical evaluation of this system is made through a deterministic smart grid simulation case study as well as through pseudo-stochastic case studies. The other major contribution of this thesis consists of designing a novel multi-armed bandit learning-based adaptive multi-agent system to tackle the effects of real-life uncertainties on the system. The evaluation of this proposed system is carried out through stochastic smart charging case studies.

To facilitate the reproducibility of the results reported in this thesis manuscript, the source code of the developed decentralized control algorithms is available here: https://gitlab.com/satie.sete/combinatorial-bandits-for-smart-grid.

Organization

The organization of this thesis manuscript is illustrated in Figure 1.11. Chapter 1 covers the three main aspects of this thesis i.e., the motivation, the contributions, and the organization. In Chapter 1, a detailed literature review of the studied smart grid problem has been presented. This chapter highlighted the advantages of applying optimization techniques to control smart grids in real-time. A discussion on different smart grid optimization strategies, categorized based on the system's architecture, utilized optimization technique(s), and application(s), was also made in Chapter 1. The novelty of this thesis along with its main contributions was also discussed in Chapter 1. The contributions of this thesis are described in Chapters 2, 3, 4, and 5. These technical chapters are based on either the adaptive multi-agent system (AMAS) theory, multi-armed bandit (MAB) theory or both. The content of these chapters is described as follows:

Chapters based on the AMAS theory

In Chapter 2, the principles of adaptive multi-agent system theory are applied to develop a decentralized smart grid energy management system. This system comprises reactive agents that respond to instantaneous environmental changes, aiming to opti-

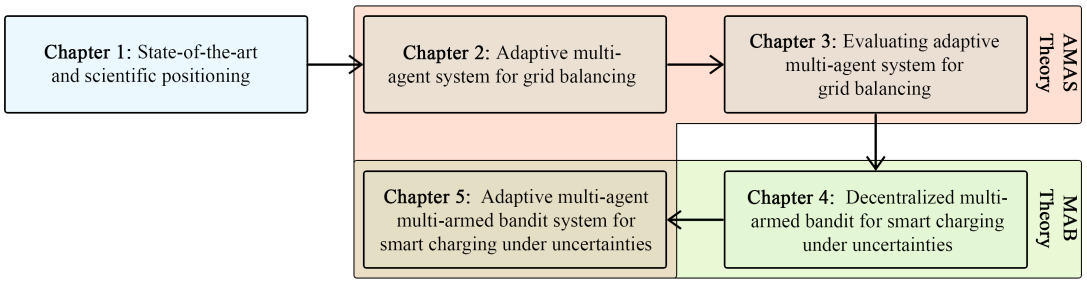

Figure 1.11: Manuscript organization.

mize energy flows within the smart grid. In Chapter 3, the designed AMAS system is thoroughly evaluated through simulation-based experiments. The performance is compared with other baseline control strategies, such as uncontrolled charging strategy and mixed-integer linear programming optimization, while considering different smart grid sizes. The evaluation includes both the quality of the solutions and the computational resources required (time and memory to obtain the desired solution) based on the number of electrical vehicles in the system. Additionally, a pseudo-stochastic smart grid study is designed to assess the system's performance under real-life uncertainties. In Chapter 5, the AMAS theory is once again employed, this time in combination with the MAB theory, to model a decentralized smart grid energy management system that effectively addresses real-life uncertainties.

Chapters based on the MAB theory

Chapter 4 introduces a decentralized energy management system designed using the principles of combinatorial multi-armed bandit theory. This chapter focuses on implementing learning capabilities to aid agents in making decisions under uncertainties. Building upon this, Chapter 5 combines the concepts of a combinatorial multi-armed bandit with adaptive multi-agent system design, leveraging the strengths of both approaches. A comprehensive evaluation of the final learning-based decentralized energy management system is also provided in Chapter 5. The optimization performance is assessed by comparing it with earlier described baseline algorithms to address the smart charging optimization problem.

The thesis culminates with a comprehensive conclusion that assesses the validity of the hypothesis proposed in this work. Additionally, the chapter explores potential short-term and long-term research directions in this area of study.

1.5 Publications

Based on the research presented in this thesis, the following publications have been made:

Peer-reviewed journal

- Zafar, S., Blavette, A., Camilleri, G., Ben Ahmed, H., and Agbodjan, J. J. P. Decentralized optimal management of a large-scale EV fleet: Optimality and computational complexity comparison between an adaptive MAS and MILP. *International Journal of Electrical Power & Energy Systems*, 147 (2023), 108861.

Peer-reviewed international conference with proceedings

- Zafar, S., Maurya, V., Blavette, A., Camilleri, G., Ben Ahmed, H., and Gleizes, M. P. Adaptive multi-agent system and mixed integer linear programming optimization comparison for grid stability and commitment mismatch in smart grids. In 2021 *IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)* (2021), pp. 01–05. Espoo, Finland (virtual). [Best paper award]

- Zafar, S., Féraud, R., Blavette, A., Camilleri, G., and Ben Ahmed, H. Decentralized Smart Charging of Large-Scale EVs using Adaptive Multi-Agent Multi-Armed Bandits. In 2023 *International Conference & Exhibition on Electricity Distribution (CIRED)* (2023). Rome, Italy.

- Zafar, S., Féraud, R., Blavette, A., Camilleri, G., and Ben Ahmed, H. Multi-Armed Bandits Learning For Optimal Decentralized Control of Electric Vehicle Charging. In 2023 *IEEE PowerTech* (2023). Belgrade, Serbia.

1.6 Conclusion

This chapter highlighted the importance of smart grid optimal control and provided the positioning of this thesis through a comparison with existing literature studies. At the beginning of this chapter, a brief discussion was made on the topic of power systems evolution. General trends, such as an increasing share of distributed energy resources and an increased focus on decentralization, were highlighted. It explained how distributed energy resources can help us in achieving our efficiency, environmental, and economic targets. However, these resources would increase the degree of stochasticity in power systems due to their variable and uncertain nature. Thus, smart grid control strategies will be an integral part of future power systems. One efficient way to tackle this variability can be through optimal control of flexible grid entities to provide flexibility to the grid. A comparison of different literature studies focusing on this topic was presented. These studies were classified into different groups based on their proposed architectures, i.e., centralized, hierarchical, or decentralized. Finally, it was stated that decentralized smart grid control is a prominent topic of research and the novelty of the proposed decentralized control system in this thesis was discussed. This chapter concluded with a detailed presentation of the contributions made by this thesis and the organization of the manuscript. In the next chapter, an in-depth design of the proposed adaptive multi-agent system for real-time energy flow optimization in large-scale smart grids is presented.

Chapter 2

Adaptive multi-agent system for grid balancing

> A whole is greater than the sum of its parts.
>
> Unknown

Summary

This chapter focuses on the design of a multi-agent system to optimize energy flows in smart grids. Specifically, the concepts of adaptive multi-agent systems are used to design the proposed multi-agent system. The developed decentralized system is intended to work in real-time and utilizes a feedback control algorithm for its decision-making. The decentralization aspect of the proposed multi-agent system makes it an excellent candidate to handle the optimization of large-scale smart grids.

Contents

The forthcoming two chapters in this thesis are dedicated to the design and evaluation of a fully decentralized smart grid control system, utilizing the concepts of adaptive multi-agent systems. The present chapter provides a comprehensive explanation of the fundamental concepts essential for understanding the proposed system, along with a detailed exposition of its design. In the subsequent chapter, Chapter 5, a comprehensive evaluation of the system will be conducted through simulation-based experiments. The primary focus of the developed adaptive multi-agent system in this chapter is to address the real-time grid balancing problem, particularly from the standpoint of a balance responsible party. The studied smart grid optimization problem is presented first in Section 2.1. Relevant work and scope of this thesis are defined in Section 2.2. Moving forward, the philosophy behind an adaptive multi-agent system (AMAS) and its underlying theory is explained in Section 2.3. Afterward, the proposed AMAS system to maintain real-time energy balance in smart grids is detailed in Section 2.4. Finally, the conclusion of this chapter is presented in Section 2.5.

2.1 Studied smart grid problem

A variety of smart grid optimization problems can be studied based on the modeled objective function. The optimization problem of controlling the instantaneous charging powers of EVs to provide energy imbalance ancillary services is studied here [200]. In this section, the required background to understand the studied problem is described first. Subsequently, the mathematical formulation of our smart grid optimization problem is presented.

Description

Existing electrical distribution networks may suffer from a variety of challenges arising due to the inclusion of new elements in the system, such as photovoltaics (PVs) and electric vehicles (EVs) [169]. These new elements would result in bi-directional power flows in an electrical system that may have been previously designed only to manage uni-directional power flows (i.e., from power stations to the consumers). More specifically, the instantaneous imbalance between the generation and consumption of electricity may arise due to the uncertain nature of PV energy production and the probabilistic nature of EV's arrival and departure times [101].

In energy markets, each transmission system operator (TSO) is obliged to maintain the balance between instantaneous production and consumption in real-time [71]. This control operation is outsourced by the TSO to the so-called *balance responsible parties* (BRPs) [105]. Each BRP has a set of loads and power sources (including exports and imports) in its perimeter, called the *balance perimeter* (BP). This model is shown in Figure 2.1. The objective of each BRP is to maintain the balance between instantaneous production and consumption in its balance perimeter. Each BRP submits a day-ahead power consumption/production schedule to the transmission system operator (TSO). This BRP schedule is defined on a sub-hourly basis (soon to be harmonized to a quarter-hourly basis in Europe), known as the *imbalance settlement period* [108]. A BRP may utilize different existing techniques to forecast the day-ahead PV energy pro-

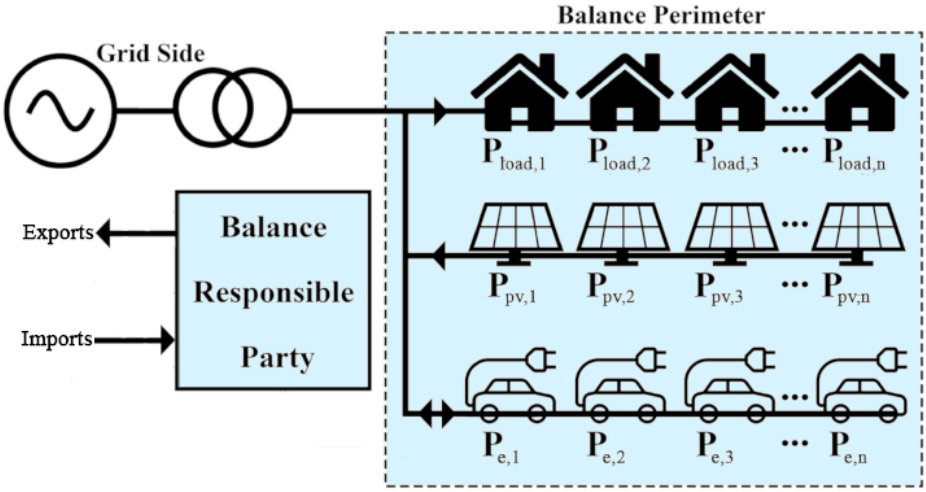

Figure 2.1: Diagram depicting a BRP and its balance perimeter.

duction and load consumption). The forecasted energy production/consumption could be used to formulate the day-ahead power consumption/production schedule submitted to the TSO. However, this day-ahead energy production/consumption forecast is prone to forecasting errors. Popular PV energy production forecasting techniques include persistence model, statistics-based models such as seasonal autoregressive integrated moving average (SARIMA), machine learning models based on artificial neural networks (ANNs), binarized neural networks (BNNs), long short-term memory (LSTM) neural networks, etc [163]. The amount of error in the forecast depends on the utilized techniques. The box plot of mean error against the studied forecasting technique is shown in Figure 2.2. This plot is obtained through the database consisting of 180 case studies from the literature on PV output forecasting [136].

Thus, due to these inherent errors in the BRP day-ahead schedule, balancing the grid in real-time becomes a complex task. An instantaneous imbalance can compromise the stability of the distribution grid [99]. Such potential problems may result in

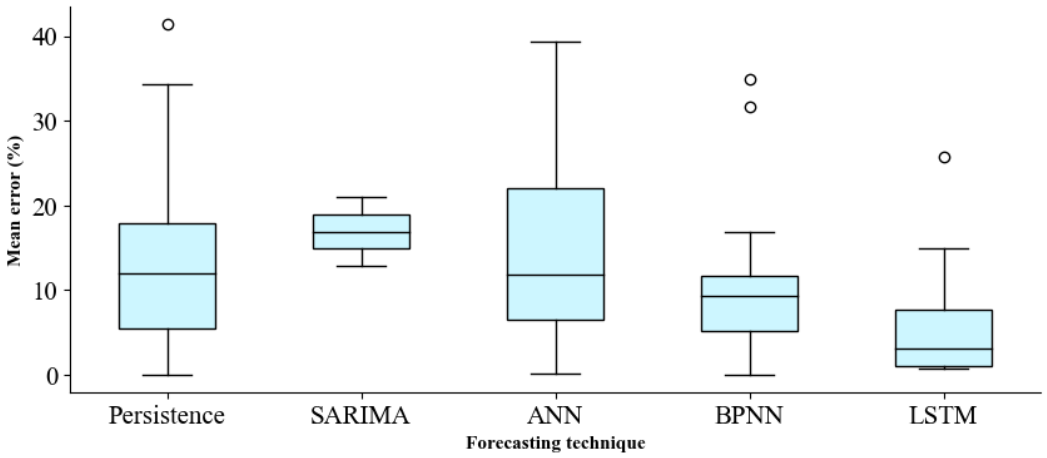

Figure 2.2: Reported mean error against day-ahead PV production forecasting techniques [136].

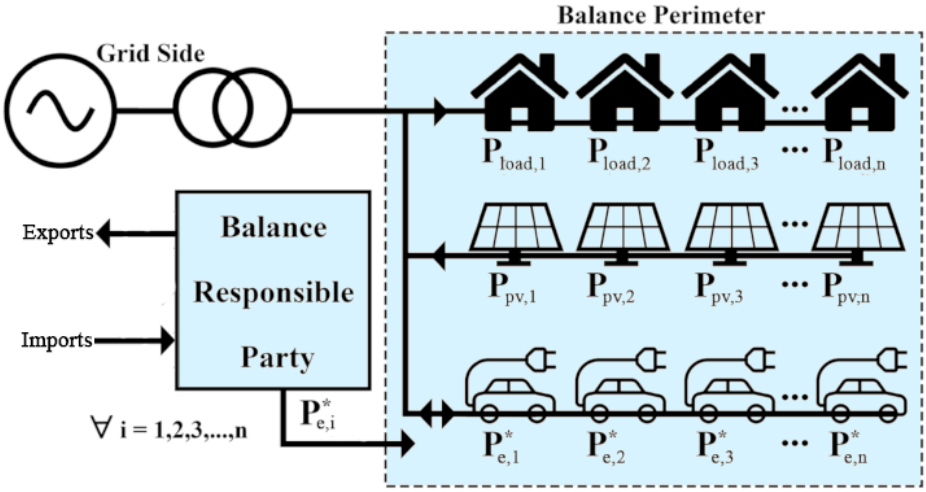

Figure 2.3: Interactions of a BRP with flexible entities in its perimeter to perform optimization.

diminishing the quality of electricity supply to consumers and may also increase the degradation of electrical equipment in the network [74]. Grid reinforcement may help in tackling the mentioned challenges. However, it may be costly and time-consuming [80]. Hence, flexible solutions may be preferred over grid reinforcement solutions.

A BRP could utilize the flexible entities (such as EVs) present in its perimeter to balance the instantaneous production/consumption in its perimeter, shown in Figure 2.3. The control can be centralized (i.e., BRP determining instantaneous charging power of all EVs), decentralized (i.e., each EV optimizing its instantaneous charging power while also communicating with its BRP), or a hybrid of both strategies. However, such control must not result in the constraint violation of other market actors (e.g., provoking congestion in the DSO network). These interactions of different market actors at different levels make this optimization problem even more complex. Centralized optimization can be deployed to tackle this optimization problem on a small-scale. However, centralized optimization methodologies suffer from inherent shortcomings (such as lack of scalability, single point of failure, data privacy concerns etc.) [3]. Thus, they may not be well suited to control a large-scale smart grid in real-time. Decentralization of the system through the philosophy of multi-agent systems (MASs) could be the way forward. It must be noted that in the current energy markets, BRPs are compensated based solely on their consumption or production, without the incorporation of any flexibility mechanisms. The solution proposed here represents a potential future avenue.

In the next subsection, the mathematical formulation of the studied smart grid optimization problem is described. Subsequently, an adaptive multi-agent system to solve the described optimization problem is presented in Section 2.4.

Mathematical Formulation

The studied optimization consists of an objective function and a set of constraints. The objective is to minimize the studied objective function while also satisfying all constraints. Descriptively, the optimization problem is summarized as follows:

- **Objective:** The objective is to minimize the instantaneous mismatch between production and consumption, which may arise due to inherent errors in a time series forecasting technique. A BRP can perform this minimization by controlling the instantaneous charging/discharging powers of each electric vehicle present in its balance perimeter.

- **Constraints:** A BRP may utilize the instantaneous charging behavior of each EV in its perimeter. However, it has to satisfy a set of constraints while doing so. The network should remain stable at all instants (i.e., distribution system operator (DSO) constraints). Furthermore, the battery of each EV should be adequately charged at its departure time. This constraint guarantees that each EV would have an adequate charge in its battery for its owner to have a smooth journey.

Objective function modeling

Let $\tilde{P}(N)$ denotes the scheduled day-ahead average production/consumption in a BRP's perimeter during the N-th imbalance settlement period. Let there be a total of N_{end} imbalance settlement periods during each day. The goal of a BRP is to control instantaneous production/consumption in its perimeter $P_{BRP}(t)$ to minimize the difference between $\tilde{P}(N)$ and $P_{BRP}(t)$ during each imbalance settlement period. Let n be the total duration of each imbalance settlement period and let Δt be the resolution (duration of each decision interval i.e., second, minute, hour, etc.) of the optimization problem. The objective function of the problem under study is defined as per Equation (2.1) [1].

Grid balancing problem's objective function

$$\min_{P_{e,a}(t)} E_{mis}(\tilde{P}(N), P_{BRP}(t)) = \min_{P_{e,a}(t)} \sum_{N=1}^{N_{end}} \left| \left(\tilde{P}(N) - \frac{\sum_{t=1}^{n} P_{BRP}(t)}{n} \right) \Delta t \right| \quad (2.1)$$

An illustration depicting the relationship between the imbalance settlement period variable N and the time-related variables t, and Δt in the studied optimization problem is given in Figure 2.4. The scheduled average production/consumption $\tilde{P}(N)$ is calculated using the forecasted time series data of PVs, EVs, and household loads present in a balance perimeter, given in Equation 2.2. Term $\tilde{P}_{p,a}(t)$ is the forecasted instantaneous PV output at bus a, $\tilde{P}_{e,a}(t)$ is the forecasted instantaneous EV load at bus a, and $\tilde{P}_{l,a}(t)$ is the forecasted instantaneous household load at bus a. Also, let $e_{p,a}(t)$ be the instantaneous error in the forecasted PV output at bus a, $e_{l,a}(t)$ be the instantaneous error in the forecasted household load profile at bus a, and $e_{e,a}(t)$ be the instantaneous error in the forecasted EV load at bus a[2]. Then, the average instantaneous produc-

[1] It should be noted that this objective function can be considered stringent from a BRP's perspective as it is assumed that any deviation from the submitted day-ahead schedule is penalized (which is not always the case).

[2] These error values of time series forecasting are also known as residual errors of the prediction. Residual errors of a prediction can be obtained by subtracting the predicted value from the real observed value.

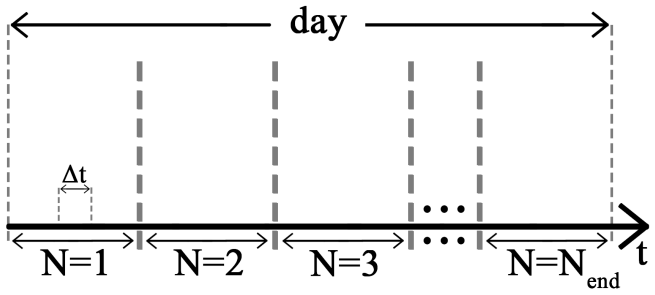

Figure 2.4: Illustration of the relationship between variables N, t, and Δt.

tion/consumption in a balance perimeter consisting of A electrical buses, during N-th imbalance settlement period is calculated using Equation (2.2).

$$\tilde{P}(N) = \frac{1}{n} \sum_{a=1}^{A} \sum_{t=1}^{n} \left(\tilde{P}_{p,a}(t)(1 + e_{p,a}(t)) - \tilde{P}_{l,a}(t)(1 + e_{l,a}(t)) - \tilde{P}_{e,a}(t)(1 + e_{e,a}(t)) \right) \tag{2.2}$$

The instantaneous BRP production/consumption $P_{BRP}(t)$ is defined as the sum of instantaneous PV output $P_{p,a}(t)$ on bus a, instantaneous household load $P_{l,a}(t)$ on bus a, and instantaneous EV load $P_{e,a}(t)$ on bus a for all A electrical buses in the network. It is described mathematically in Equation (2.3).

$$P_{BRP}(t) = \sum_{a=1}^{A} \left(P_{p,a}(t) - P_{e,a}(t) - P_{l,a}(t) \right) \tag{2.3}$$

Terms $P_{p,a}(t)$ and $P_{l,a}(t)$ in the studied optimization problem are always non-negative real numbers and represent instantaneous production through photovoltaics and instantaneous consumption by the household load respectively. The instantaneous EV load $P_{e,a}(t)$ can either be positive, negative, or zero. This is summarized as follows:

$$P_{e,a}(t) \begin{cases} > 0 & \text{if the EV is charging} \\ < 0 & \text{if the EV is discharging} \\ = 0 & \text{if the EV is neither charging nor discharging} \end{cases} \tag{2.4}$$

Constraints' modeling

As stated earlier, each BRP must satisfy a set of constraints while minimizing its objective function. This set of constraints includes the constraints of distribution system operators (DSOs) and the constraints of prosumers. The physical power flow constraints of the distribution network must also be satisfied. Power flows in a distribution network must abide the Ohm's law [67]. Let $P_a(t)$ be the instantaneous active power at bus a. This instantaneous active power is equal to the difference between total generated $P_{a,gen}(t)$ and total demanded active power $P_{a,dem}(t)$ at bus a. Similarly, the

instantaneous reactive power at bus a is equal to the difference between total generated $Q_{a,gen}(t)$ and total demanded reactive power $Q_{a,dem}(t)$ at bus a. The set of distribution network's physical constraints is given in Equations (2.5)-(2.9).

Distribution network's physical constraints

$$P_a(t) = P_{a,gen}(t) - P_{a,dem}(t) \tag{2.5}$$

$$Q_a(t) = Q_{a,gen}(t) - Q_{a,dem}(t) \tag{2.6}$$

$$\sum_b P_{ab}(t) = P_a(t) \tag{2.7}$$

$$\sum_b Q_{ab}(t) = Q_a(t) \tag{2.8}$$

$$P_{ab}(t) + iQ_{ab}(t) = V_a(t)\left(V_a^*(t) - V_b^*(t)\right) Y_{ab}^* \tag{2.9}$$

Equations (2.7) and (2.9) state that the inflow of powers to the bus is equal to the outflow of powers at each bus a. Equation (2.9) relates root-mean-square voltages at bus a and bus b (i.e., $V_a(t)$ and $V_b(t)$ respectively) with admittance matrix Y_{ab}^* of electrical line between bus a and bus b. The constraints of a distribution system operator (DSO) must be satisfied as well. Each DSO must keep its distribution network stable i.e., there should not be any electrical current congestion or voltage limits violation in its distribution network [107], [153]. These constraints are stated in Equations (2.10)-(2.13).

Distribution network operator's constraints

$$I_{ab}(t) < I_{ab,max} \tag{2.10}$$

$$V_{a,min} < |V_a(t)| < V_{a,max} \tag{2.11}$$

$$P_a(t) < P_{a,max}(t) \tag{2.12}$$

$$Q_a(t) < Q_{a,max}(t) \tag{2.13}$$

Equation (2.10) states that the root-mean-square electrical current flowing through the electrical line connecting bus a and bus b should be lower than its rated value $I_{ab,max}$. The magnitude of the instantaneous root-mean-square voltage at each bus $V_a(t)$ must also remain between a maximum value $V_{a,max}$, and a minimum value $V_{a,min}$, as shown in Equation (2.11). Distribution system operators (DSOs) may also put a limit on the instantaneous power drawn at each bus a, given in Equations (2.12) & (2.13).

The instantaneous charging power of an EV $P_{e,a}(t) \in [P_{e,a,min}, P_{e,a,max}]$, at bus a is defined as the sum of the forecasted instantaneous EV charging power $\tilde{P}_{e,a}(t)$ and the term $\Delta P_{e,a}(t)$, which is the decision variable of the studied real-time optimization

problem. This relationship is given in Equation (2.14). At its departure time $t_{e,a,depart}$, electric vehicle (EV) e, connected to electrical bus a, should have a desired minimum state of charge $SoC_{e,a,depart}$. This constraint would allow each EV owner to have a smooth journey. The state of charge (SoC) of a battery is the level of charge of an electric battery relative to its capacity. It is generally given as a percentage value. The instantaneous SoC value of a battery is related to its past instant's SoC value, its capacity $E_{e,a,bat}$, its charging/discharging efficiency $\eta_{e,a}$, and its instantaneous charging/discharging power $P_{e,a}(t)$. To limit the rate of battery degradation, the instantaneous state of charge (SoC) of each EV must remain within certain bounds[55]. The SoC of each EV $SoC_{e,a}(t)$ should be between a set maximum SoC value $SoC_{e,a,max}$, and a set minimum SoC value $SoC_{e,a,min}$. An electric vehicle's state of health (SoH) must also be greater than zero. The battery's SoH variable makes it possible to estimate how much it has deteriorated over time. Its values range from 0 to 1. According to the Equation (2.16), a battery's end of life is indicated by a 20% capacity fade. When an EV battery reaches the end of its useful life, it must be replaced. The term $E_{e,a,tp}$ refers to an EV battery's energy throughput. It is the total amount of energy a battery can store and release over the course of its lifetime. A battery's capacity, efficiency, cycle life, and depth of discharge all affect its throughput. These constraints are given as follows:

Prosumer's constraints

$$P_{e,a}(t) = \tilde{P}_{e,a}(t) + \Delta P_{e,a}(t) \quad \text{s.t.} \quad P_{e,a}(t) \in [P_{e,a,min}, P_{e,a,max}] \tag{2.14}$$

$$SoC_{e,a,min} < SoC_{e,a}(t) = SoC_{e,a}(t-1) + \frac{P_{e,a}(t)\eta_{e,a}\Delta t}{E_{e,a,bat}} < SoC_{e,a,max} \tag{2.15}$$

$$SoH_{e,a}(t) = SoH_{e,a}(t-1) - \frac{P_{e,a}(t)\Delta t}{0.2E_{e,a,tp}} > 0 \tag{2.16}$$

$$SoC_{e,a}(t_{e,a,depart}) > SoC_{e,a,depart} \tag{2.17}$$

2.2 Relevant research and scope

Related work

An EV energy management system to maintain demand/supply balance has been proposed in [83]. However, DSO constraints have not been included in the problem formulation of [83]. In [196], another energy management system to minimize power imbalances by controlling electric vehicles' charging/discharging has been presented. In this system, no DSO constraints have been modeled as well. There exist EV energy management solutions that take both prosumers and DSO constraints into consideration, such as [197] and [14]. A coordinated charging methodology to minimize the impact of mismatches due to uncertainties by controlling the charging power of each EV is proposed in [197]. Both prosumers and DSO constraints have been considered in

this designed system. In [14], an EV charging management algorithm to minimize energy losses has been proposed. Constraints of both DSO and prosumers have also been modeled in this system. However, the architecture of these systems (i.e., [83], [196], [197] and [14]) are centralized in nature. Thus, these systems may not be scalable and may not be able to manage large-scale smart grids in real-time.

The use of electric vehicles to provide support to balance responsible parties has also been studied in [125] and [186]. In [125], the aim was to tackle the uncertainty in wind energy production by balancing the grid in real-time utilizing electric vehicles. Charging constraints of electric vehicles have been considered. However, the DSO constraints were not included in the optimization formulation. On the other hand, the objective was to maintain the balance between energy suppliers and consumers through the utilization of electric vehicles in [186]. The proposed control architectures of both systems ([125] and [186]) are not fully decentralized and thus may suffer from drawbacks of centralization due to their hierarchical system architectures.

To mitigate the drawbacks of centralization, a decentralized energy management system has been proposed in [182]. The desired goal of this system is to minimize the imbalance cost of a balance responsible party. This is achieved by controlling the instantaneous charging/discharging of electric vehicles. The transformer congestion constraint has been considered but the voltage stability constraint has not been included in [182]. Whereas the proposed system in [137] considers voltage stability constraint as well to achieve the same objective of grid balancing. The simulation case study performed is not large-scale.

Scope

The scope of the studied energy balancing problem is defined in Figure 2.5. As smart grid control is a complex subject, thus, defining the scope becomes imperative. By defining the scope, the area covered by this study, the boundaries, and the limitations are clearly described. This would help in the system design stage, the evaluation of the developed system, and highlighting the most important points of this study.

It can be clearly observed in Figure 2.5 that the emphasis is given to the architecture, scale, control's temporal resolution, considered constraints, EV's charging technology, and comparisons made in this study. The *architecture* of the energy management system is selected to be decentralized to eliminate the disadvantages of centralization and hybrid architectures. As the designed system is intended to manage real-life smart grids, thus, the *scale* is set to large. It should be noted that in this thesis, small-scale refers to an electrical grid with less than 100 EVs, while large-scale means a smart grid with more than 10,000 EVs. An electrical grid with EVs between 100 and 10,000 is considered a medium-scale grid. *Control* can be day-ahead (scheduling), near real-time (hours), or real-time (minutes or seconds). In this study, it is intended that the control of the developed system should be in real-time to maintain the stability of the system. Both prosumers and DSOs would be among the biggest stakeholders of the future distribution networks. Thus *constraints* of both of these market actors are considered here. In this study, both grid-to-vehicle (G2V) and vehicle-to-grid (V2G) controlled *charging technologies* are considered. The V2G technology can especially be useful to help a BRP in case of an over-consumption challenge. Finally, to evaluate

the performance of the developed system a *comparison* is made with baseline strategies in terms of optimality as well as in terms of the required computation time and memory.

As evident in Figure 2.5, a relatively lower focus is given to evaluation, DERs, stochasticity, and forecaster aspects in this study. To carry out an *evaluation* of the developed system simulation case studies are performed. In terms of *DERs* consideration, only penetration of PVs is considered on the distribution side in this study. *Simulation* studies are a natural first step before deploying a smart grid control solution as one can evaluate the performance of the developed control system under a range of different operating conditions without risking the electrical grid's infrastructure. In terms of *stochasticity*, only pseudo-stochasticity is studied here. It is the first step towards complete stochasticity. If the proposed system does not function efficiently under pseudo-stochasticity (hence under stochasticity as well), advanced techniques such as machine learning can be utilized to minimize the effects of uncertainties on the system. Finally, a BRP requires a *forecaster* to calculate its day-ahead schedule. A simple persistence model is used here to forecast time-series data for the next day.

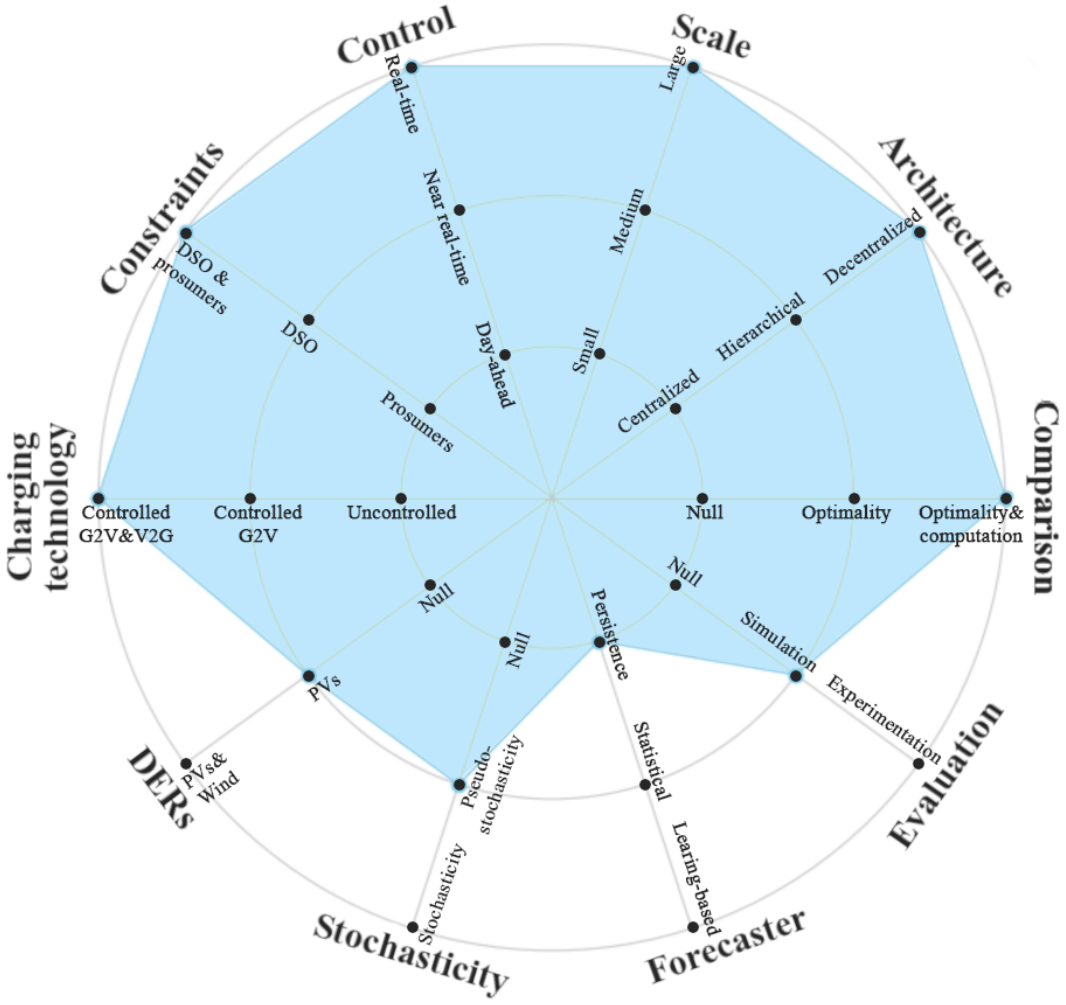

Figure 2.5: Scope of the studied smart grid problem in this chapter.

Here, the focus is not on designing a novel time-series forecaster. Instead, the focus is on minimizing energy imbalances in real-time that may occur due to inherent errors in any time-series forecaster.

2.3 Introduction to adaptive multi-agent systems

The studied smart grid optimization problem can be solved in a centralized way as well as in a decentralized manner. A visual representation of both optimization architectures (centralized and decentralized) is given in Figure 2.6. In contrast to centralized optimization algorithms, a decentralized optimization algorithm does not need to gather information at a single node to perform optimization. Instead, all the entities present in the system may communicate with each other to optimize the system [118]. According to [54], a *multi-agent system* (MAS) can naturally be designed to be a decentralized system. To fully understand the functioning of a multi-agent system, some basic definitions are presented as follows:

Multi-agent system definitions

Definition 2.3.1 (Multi-agent system)**.** A multi-agent system is a physical or software system consisting of a number of agents interacting to achieve (a) desired goal(s).

Definition 2.3.2 (Agent)**.** An agent is a physical or software entity in a multi-agent system that perceives its environment and acts over it.

Definition 2.3.3 (Environment)**.** Everything outside an agent and with which an agent may interact is termed that agent's environment.

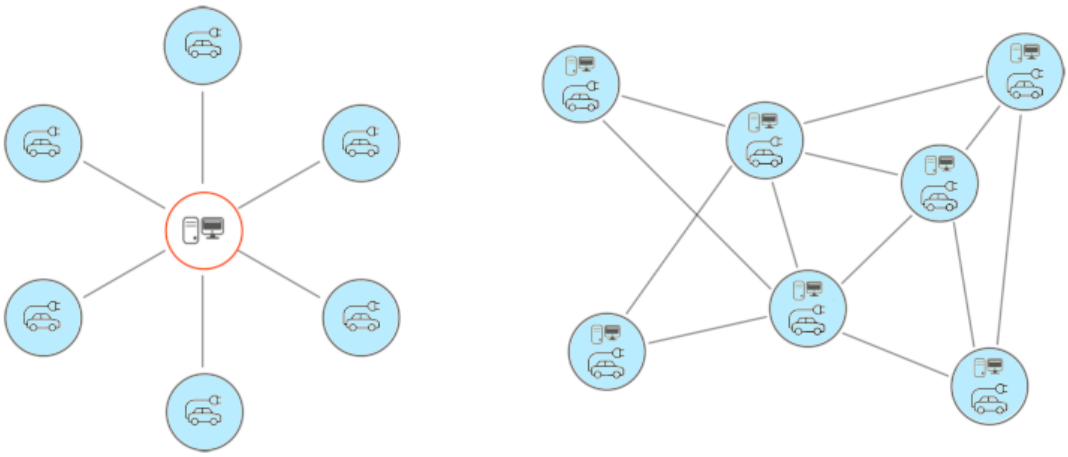

Figure 2.6: Comparison of centralized (left) and decentralized (right) architectures for smart grid optimization.

A multi-agent system consists of agents that interact with each other and their environment to achieve a desired goal(s). These agents may hold several important characteristics such as [49]:

- **Autonomy:** An agent may be autonomous i.e., it can control its behavior independently and it does not rely on any other entity for its operation.
- **Sociability:** An agent may be social i.e., it can communicate with other agents to carry out its desired functionality.
- **Proactivity:** An agent may be proactive i.e., it can take the initiative and opportunistically adopt new goals.
- **Reactivity:** An agent may be reactive i.e., it can respond and act in a timely fashion to the changes in its environment.
- **Locality:** An agent may be percept only locally i.e., it can observe only a portion of the whole system as its environment.
- **Interactivity:** An agent may be interactive i.e., it can interact (cooperatively or non-cooperatively) with other agents in the system to maximize utility.
- **Learning:** An agent may be learn i.e., it can learn to maximize its utility by keeping a history of its past interactions with the environment.

An agent can be modeled to have any of the above-mentioned properties. Multi-agent systems have found a range of real-life applications such as, robotics [102], transportation [170], data analytics [20], and power systems [147].

There exists a specific sub-class of multi-agent systems i.e., adaptive multi-agent systems. The engine of adaptive multi-agent system (AMAS) theory is *cooperation*. An AMAS focuses on achieving the desired objective through cooperative actions among its agents. Cooperation among agents can be defined as the act of working towards a common objective or some underlying benefit. Agents in an AMAS cooperate to achieve a common global goal of the system. A cooperative attitude among all agents of an AMAS would lead to the satisfaction of the following properties:

- **Sincerity:** Each agent is sincere with all other agents in the system.
- **Prosociality:** Each agent is ready to help, when it is possible, another agent facing a more difficult situation.
- **Reciprocity:** Each agent satisfying the above-given properties knows that these properties will also be satisfied by all other agents of its system.

The theorem of functional adequacy has been presented in [21] to demonstrate the improvements coming from cooperation in an AMAS. This theorem is given below:

Functional adequacy of a system can be defined as its ability to meet the desired requirements by performing the intended tasks, and delivering the desired outcomes. Instead of designing a complex system that satisfies the required functionality as a

Theorem of functional adequacy

Theorem 2.3.1 (Functional adequacy). For any functionally adequate system, there is at least one cooperative internal medium system that fulfills an equivalent function in the same environment.

whole, one can focus on designing a simpler cooperative internal system utilizing the AMAS theory. The adequacy of this simpler cooperative internal system is guaranteed by the theorem of functional adequacy. The definition of a cooperative internal system along with the situations when a system cannot be classified as a cooperative internal medium system is given below:

Cooperative internal system

Definition 2.3.4 (Cooperative internal system)**.** A cooperative internal medium system is a system where no non-cooperative situation (NCS) exists.

Definition 2.3.5 (Non-cooperative situation)**.** An AMAS agent is said to be in a non-cooperative situation when:

- a perceived signal is not understood by it, or the signal is ambiguous;
- the perceived information does not result in any activity process;
- the action of an agent is not useful for its system.

The requirement for a cooperative internal medium system is to avoid any non-cooperative situation (NCS). Thus, the AMAS objective is to design a system in which agents undergo cooperative interactions with each other to prevent any non-cooperative situation from arising or to manage it if any non-cooperative situation arises. This can be through a criticality value for each agent. Criticality can be defined as the local measure of the dissatisfaction degree of an agent. All agents can interact cooperatively with each other. Each AMAS agent is cooperative to tackle any non-cooperative situation in the system. When an agent is cooperating with another agent, it helps the other agent by working towards its goal. However, the self-objective of this agent remains unchanged. This behavior of continuous re-organization of goals shown by an AMAS agent results in self-organization. *Self-organization* has been defined as, "*the mechanism or the process enabling a system to change its organization without explicit external command during its execution time*" [49]. Based on the existence of self-organization, one can further divide multi-agent systems into two categories:

- **Strong self-organizing system:** A system with no explicit central control, either internal or external.
- **Weak self-organizing system:** A system in which re-organization may be performed under internal (central) control or planning.

An AMAS system generally falls under the category of a strong self-organizing system. It demonstrates the ability of self-organization through the earlier discussed cooperation mechanism. The main advantage of self-organization in AMAS is that one does not need to program the global functionality of a complex system within the agent. Instead, only the local objective of each AMAS agent is defined while designing a cooperative internal system, and the desired global functionality is obtained through self-organization. Cooperation in an adaptive multi-agent system can also be combined with other techniques that may lead to self-organization such as, bio-inspired methodologies (stigmergy, reinforcement learning, etc.), social-based approaches (trust-based, social functions, auction, etc.), and artificial approaches (authentication chains, tag-based models, and so on) [49]. Agents in a multi-agent system can be divided into different categories based on their design and the type of self-organization mechanism utilized. For example, in Figure 2.7, Nwana has utilized the existence of autonomy, cooperation (interactivity), and learning in agents to divide them into four main categories i.e., collaborative agents, interface agents, collaborative learning agents, and intelligent agents [138].

The self-organizing nature of an AMAS results in its adaptability [21]. In multi-agent systems, *adaptability* refers to a system's ability to modify its behavior in response to changes in the environment. Adaptability is particularly important to tackle dynamic nonlinear environments. A *dynamic nonlinear environment* is a complex environment that involves unpredictability due to non-linear complex relationships among different system variables. A classical multi-agent system based on a fixed set of rules may not be able to fully optimize a complex system due to the unpredictable and dynamic nature of the system's environment. On the other hand, adaptive multi-agent systems are designed to be flexible and thus adaptable to changes in their environments. An electrical network can be classified as a dynamic nonlinear envi-

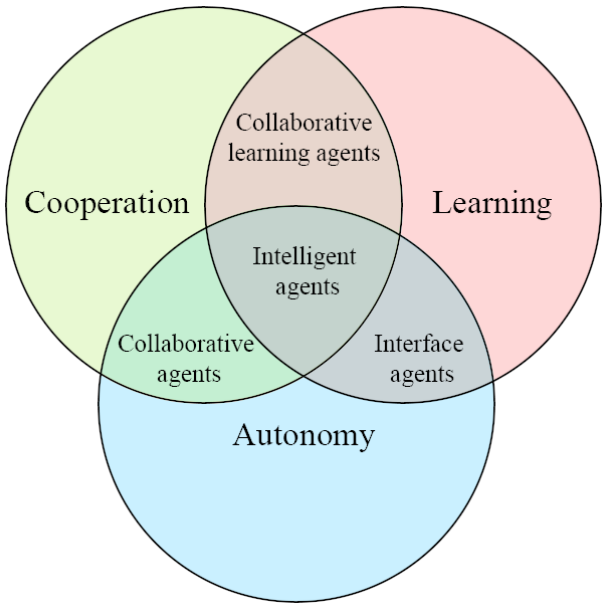

Figure 2.7: Agent types proposed by Nwana based on autonomy, cooperation, and learning in agents [138].

ronment as it is constantly evolving and it involves complex interactions between different system elements. Consequently, adaptive multi-agent systems are better suited to optimize real-time energy flows in a smart grid because they offer flexibility and adaptability, which would increase a smart grid's efficiency, reliability, and resilience. Furthermore, adaptive multi-agent systems are also known to be robust, and scalable.

A key characteristic of an adaptive multi-agent system is that each agent perceives only a local view as its environment i.e., it only communicates in its *neighborhood* [21].

Adaptive multi-agent system definition

Definition 2.3.6 (Neighborhood)**.** The neighborhood of an agent, in an adaptive multi-agent system, is defined as the set of agents with which it directly interacts cooperatively to achieve its goal(s).

The difference between potential interactions in a classical fully connected multi-agent system and an adaptive multi-agent system is shown in Figure 2.8. It can be seen that in a fully connected MAS, an agent (highlighted in blue) may directly interact with all other agents (highlighted in green) in the system. However, generally, in an adaptive multi-agent system, this agent (highlighted in blue) only interacts with agents present in its defined neighborhood (highlighted in green), and does not interact with agents outside its neighborhood (not highlighted). Each AMAS agent does not know about the global objective(s) of the system. Instead, it is only cooperatively interacting with its neighboring agents. The definition of each agent's neighborhood depends on the designer of the adaptive multi-agent system. All agents are autonomous and cooperative in an adaptive multi-agent system. No single agent has an understanding of the overall objective of the system. The cooperative interactions of agents with each other help to satisfy the overall objective(s) of their system. This type of problem-solving methodology is commonly labeled as *emergent problem solving* methodology [21]. A number of adaptive multi-agent systems have been developed to tackle various real-life challenges such as big data analytics [20], supply chain optimization [68], and smart grid optimization [200].

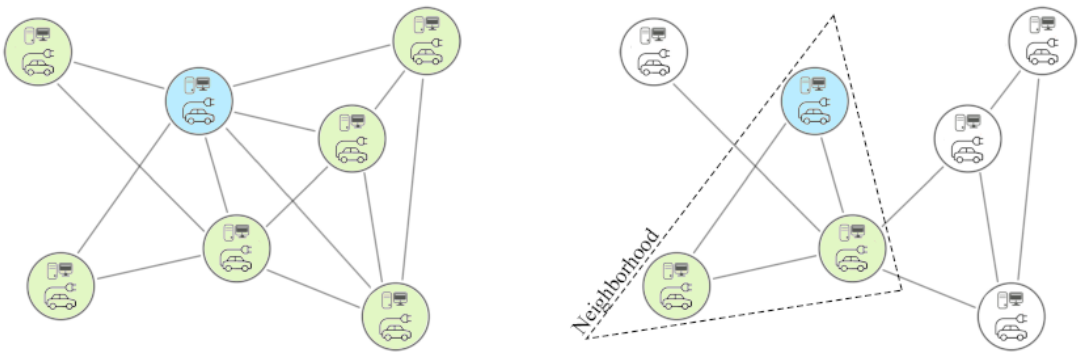

Figure 2.8: Comparison of interactions in a classical fully connected MAS (left) against an AMAS (right).

In this section, the general concepts of multi-agent systems have been defined, along with the distinctions between an adaptive multi-agent system and a more general multi-agent system. The potential advantages of applying an adaptive multi-agent system to manage a smart grid over a classical fully connected multi-agent system are also highlighted in this section. In the next section, the design of an adaptive multi-agent system designed specifically to optimize energy flows in a smart grid is discussed in detail.

2.4 Proposed adaptive multi-agent system

The proposed adaptive multi-agent system (AMAS) to optimize the smart grid problem described in Section 2.1 is presented here. The desired AMAS is a software system, while an electrical distribution network is a physical system. Thus, it is natural to map the desired physical elements present in an electrical distribution network to software agents. This mapping of physical elements to software agents is termed the *agentification* process. The software model obtained through this agentification process is called an *agentified* model, for which an example is shown in Figure 2.9. There are four types of agents in the proposed adaptive multi-agent system. These agent types are as follows:

- **Line agents:** Electrical lines present in a distribution network are modeled as line agents in the designed AMAS. The goal of each line agent is to ensure that its instantaneous electrical current does not exceed its rated electrical current value, while it also helps other agents in the system (explained in the next subsection). A line agent can ask for cooperation from electric vehicle agents to achieve its objective.

- **Bus agents:** Each electrical bus in a distribution network is modeled as a bus agent. The goal of each of these bus agents is to keep its instantaneous voltage

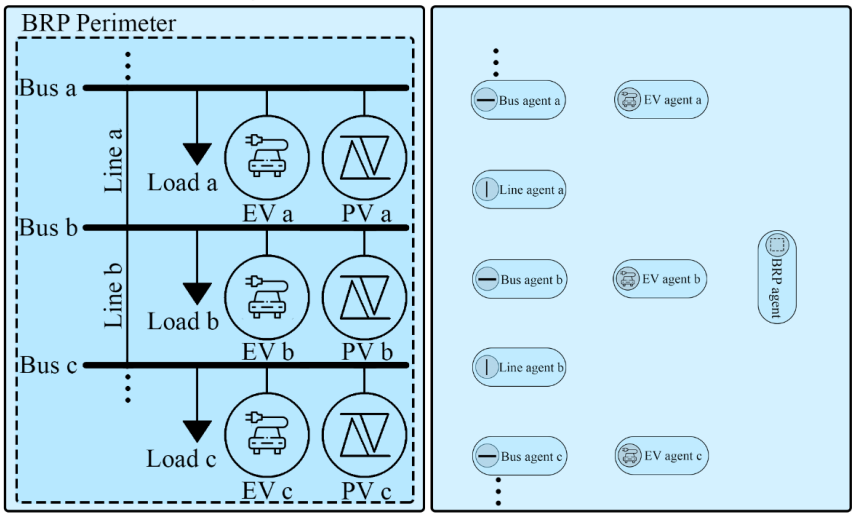

Figure 2.9: Section of a distribution network (left) and its agentified model (right).

magnitude within a desired range, while helping other agents. Each bus agent can cooperate with electric vehicle agents to achieve its desired goal.

- **Balance responsible party agents:** Balance responsible parties (BRPs) are modeled as balance responsible party (BRP) agents. The objective of each BRP agent is to minimize the potential real-time energy mismatches in its balance perimeter i.e., Equation (2.1). A BRP agent can ask for cooperation from electric vehicle agents corresponding to electric vehicles in its balance perimeter to achieve its objective.

- **Electric vehicle agents:** Electrical vehicles present in a distribution network are modeled as electric vehicle (EV) agents in the proposed adaptive multi-agent system. Each EV agent tries to satisfy the constraints of its prosumer, Equations $(2.14) - (2.17)$, while helping other agents in the system.

The agentified adaptive multi-agent system of a section of an electrical distribution network is shown in Figure 2.9. The interactions (communications) among these software agents depend on the definition of each agent's neighborhood.

Agents modeling

In this subsection, the detailed functionality of each agent type and its neighborhood are presented. The designed adaptive multi-agent system executes in a loop. During each iteration, each agent in the system tries to satisfy its objective while helping its neighboring agents. Thus, each agent faces a potential dilemma (whether to help itself or any of the neighboring agents) during each iteration of the system. To handle this dilemma, each agent is designed to hold an instantaneous criticality value (between 0 and 1). A criticality value of an agent is defined as the local measure of the dissatisfaction degree of an agent. Thus, the dilemma is tackled through the *comparison of criticalities principle*.

Comparison of criticalities principle

Definition 2.4.1 (Comparison of criticalities principle)**.** According to this principle, an agent compares its instantaneous criticality with the instantaneous criticalities of its neighboring agents. Then, the instantaneous action made by this agent is to help the agent with the highest instantaneous criticality in its neighborhood including its own.

Each agent goes through three stages during each of its cycle (iteration) [21]:

- **Perception:** In this stage, an agent gathers data from its defined environment.

- **Decision:** Based on the observed data, an intelligent decision is made.

- **Action:** In this stage, the previously selected decision is implemented by taking the required actions.

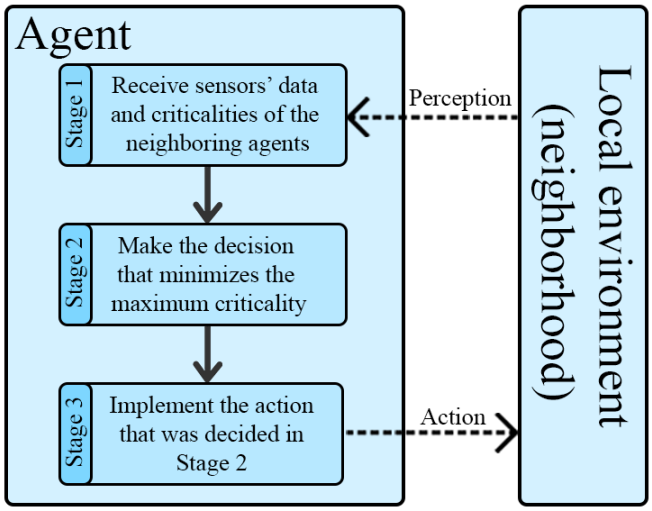

Figure 2.10: Three stages of an AMAS agent every cycle (perception, decision, and action).

The cycle of an AMAS agent during each iteration is shown in 2.10. Each agent in the designed AMAS belong to the *collaborative* class of agents according to the Nwana's agent typology, shown in Figure 2.7.

Line agent

The objective of each line agent in the system is to restrict the magnitude of electric current flowing through its corresponding electrical line below a given rated electric current value. A line agent corresponding to the electrical line connecting bus a and bus b, calculates its instantaneous criticality $Cr_{l,ab}(t)$ according the following *line agent's criticality model*:

Line agent's criticality model

$$Cr_{l,ab}(t) = \begin{cases} 0 & \text{if } I_{ab}(t) < I_{ab,th} \\ \frac{I_{ab}(t) - I_{ab,th}}{I_{ab,max} - I_{ab,th}} & \text{if } I_{max} \geq I_{ab}(t) \geq I_{ab,th} \\ 1 & \text{if } I_{ab}(t) > I_{ab,max} \end{cases} \tag{2.18}$$

The determination of the current congestion issue is indeed performed based on a comparison test between the maximum allowed flow (in both directions) and the measured flow (measured with a directional sensor). In Equation (2.18), $I_{ab}(t)$ is the instantaneous root-mean-square electrical current flowing from bus a to bus b, $I_{ab,max}$ is the rated electrical current value through the electrical line connecting bus a and bus b, and $I_{ab,th}$ is a set threshold value on the electrical current between bus a and bus b. The line criticality value $Cr_{l,ab}(t)$ is zero when the instantaneous line current is below the set threshold value, and starts increasing linearly otherwise. The line criticality of an electrical line becomes maximum (i.e., 1) when its instantaneous electrical current reaches its rated value. The relationship between the line criticality of a line agent and its instantaneous electrical current is shown in Figure 2.11.

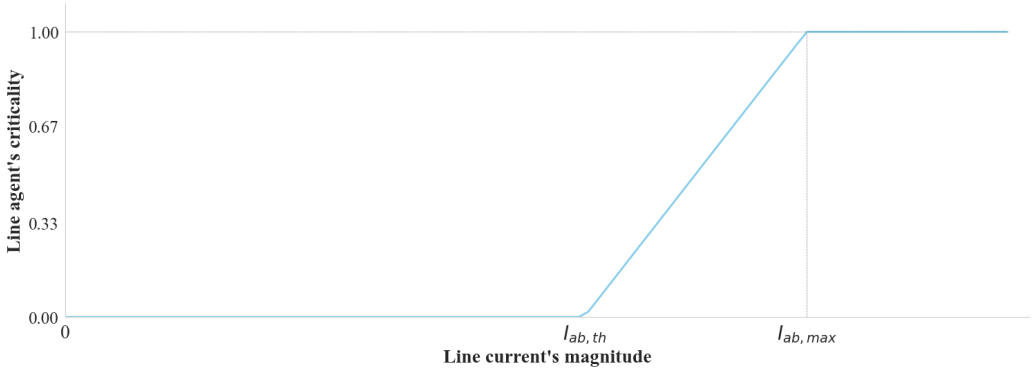

Figure 2.11: Relationship between line criticality and electrical current's magnitude.

The main objective of a line agent is to keep its criticality value equal to zero (i.e., to keep the electrical line uncongested) at all instants. In case of a non-zero criticality value (i.e., the line is congested or near congestion), it can request EV agents for assistance. As a result of EVs' cooperation, the line criticality (and the instantaneous line current) would be reduced. However, the system may immediately return to its previous congested state as soon as the earlier congested line agent stops asking for cooperation from EVs. This happens because each designed line agent includes a feedback loop (i.e., criticality dictates the instantaneous action $\longrightarrow$ action impacts the instantaneous line current $\longrightarrow$ line current determines the next instant's line criticality). This is evident in Equation (2.18), as the instantaneous criticality of a line agent depends only on the instantaneous electrical current. Therefore, instability in the form of large oscillations may occur due to the agent's continuous switching between highly critical (i.e., $Cr_{l,ab}(t) > 0$) and non-critical states (i.e., $Cr_{l,ab}(t) = 0$). These oscillations are undesirable, especially in a highly reactive system. To tackle this challenge, once a line becomes congested (i.e., $Cr_{l,ab}(t) > 0$), the congested line agent does not utilize the earlier stated line agent's criticality model (Equation (2.18)) to calculate its instantaneous criticality (which it is going to forward to its neighboring agents). Instead, it uses a memory-based model to calculate its criticality. This *memory-based line criticality model* is given as follows:

Memory-based line agent's criticality model

$$Cr_{m,l,ab}(t) = k_l Cr_{l,ab}(t) + (1 - k_l) Cr_{m,l,ab}(t-1) \tag{2.19}$$

In the given memory-based criticality model (which a line agent uses to calculate its instantaneous criticality once it becomes congested), it can be seen that the memory-based instantaneous criticality $Cr_{m,l,ab}(t)$ depends on both the instantaneous value of the electrical current ($Cr_{l,ab}(t) \propto I_{ab}(t)$), and the past history of itself $Cr_{m,l,ab}(t-1)$. Term k_l is the tuning parameter of the memory-based line criticality model. The memory-based line criticality $Cr_{m,l,ab}(t)$ converges towards the simple line criticality value $Cr_{l,ab}(t)$, if $Cr_{l,ab}(t)$ remains stationary. The rate of its convergence depends on the tuning parameter k_l. Larger values of this tuning parameter k_l result in quicker

convergence but might increase the system oscillations. Smaller values of k_l ensure none or minimal oscillations but at the cost of slower convergence.

In case of electrical congestion, a line agent can request EV agents for cooperation. However, the interactions between a line agent and an EV agent are not direct in the designed system. In an AMAS, an agent can only communicate with its neighboring agents. Furthermore, electrical cables are connected to electrical buses in a physical distribution network. Thus, the neighborhood of a line agent, that connects electrical bus *a* and electrical bus *b*, consists of bus agents corresponding to electrical buses *a* and *b*. The visual representation of a line agent's neighborhood is shown in Figure 2.12. In Figure 2.12, the neighborhood of line agent *a* is shown i.e., neighborhood *a*. The shown neighborhood includes bus agent *a* and bus agent *b*. Thus, line agent *a* can only communicate with bus agent *a* and *b* in Figure 2.12. If a line agent requires cooperative action(s) from EVs, it will send its request to its neighboring bus agents. Bus agents can include EV agents in their neighborhood (elaborated in the following subsection). Thus, the request of a congested line agent will reach EV agent(s) through bus agent(s).

Note 2.4.1

The communication happens only locally i.e., in an agent's neighborhood. For example, if line agent *a* in Figure 2.12 wants its request to reach bus agent *c*, then it will communicate its request to bus agent *b*. Afterward, bus agent *b* will pass this request to line agent *b*, which will eventually communicate the request generated by line agent *a* to bus agent *c* in Figure 2.12.

Note 2.4.2

The information flow (communication direction of the generated request) is in a single direction only i.e., only upstream or downstream. For example, if line agent *b* in Figure 2.12 receives a request from bus agent *b*, then this request may only be transferred to bus agent *c* (and it will not be transferred back to bus agent *b*).

Thus, in the proposed AMAS, if a line agent requires cooperative action(s) from EVs, it will send its request to its neighboring bus agents. Bus agents can include EV agents in their neighborhoods (elaborated in the following subsection). Thus, the request of a congested line agent will eventually reach EV agent(s) through bus agent(s). The two-way communication between bus agents and line agents follows a communication message format. According to this message format, the sent message (request) consists of an ordered pair. Both elements of this ordered pair are defined below:

- **Criticality:** The first term is the calculated criticality value. This value ranges between 0 and 1.

- **Issue:** Electrical current congestion can occur due to excessive power imports from the grid and large power exports to the grid. The desired response of flexible grid elements (EVs) depends on the type of electrical current congestion (i.e., decrease charging power when power import from the grid is high and increase charging power if power export to the grid is high). Thus, this second element of the message dual pair carries information about the critical agent's challenge.

Along with satisfying its objective, each line agent should also cooperate with its neighboring agents at any given instant. A critical agent (i.e., $Cr_{l,ab}(t) > 0$) may receive a request from another critical agent in its neighborhood. For example, a critical line agent *b* in Figure 2.12 can receive a request generated by a critical line agent *a*, through bus agent *b*. In that particular scenario, line agent *b* will have to decide if it will transfer its own criticality request to bus agent *c*, or it will forward the received criticality request of line agent *a* to bus agent *c*. This decision is made by the agent through the earlier-stated *comparison of criticalities* principle i.e., the line agent will adapt and thus forward the request with the highest criticality to bus agent *c* in Figure 2.12. Thus, a line agent cooperates with its neighboring agent by giving a higher priority in case any of its neighboring agents is more critical. The detailed functionality of a line agent is described in Algorithm 2.1.

A line agent in Algorithm 2.1 goes through its three AMAS cycle stages (perception, decision, and action). During the *perception stage*, it observed the instantaneous value of electrical current $I_{ab}(t)$ flowing through its corresponding electrical line. Based on this instantaneous electrical current value, a line agent calculates its instantaneous criticality $Cr(t)$ using either Equation (2.18), or Equation (2.19). If this line agent was not previously congested then Equation (2.18) is used to calculate its instantaneous criticality. Otherwise, Equation (2.19) is used to calculate $Cr(t)$. A line agent also calculates term $\mathcal{I}(t)$. This term holds information regarding the current issue a line agent faces. It can be **inflow**: electrical line congestion occurring due to large inflow of power from the grid; **outflow**: electrical line congestion due to large

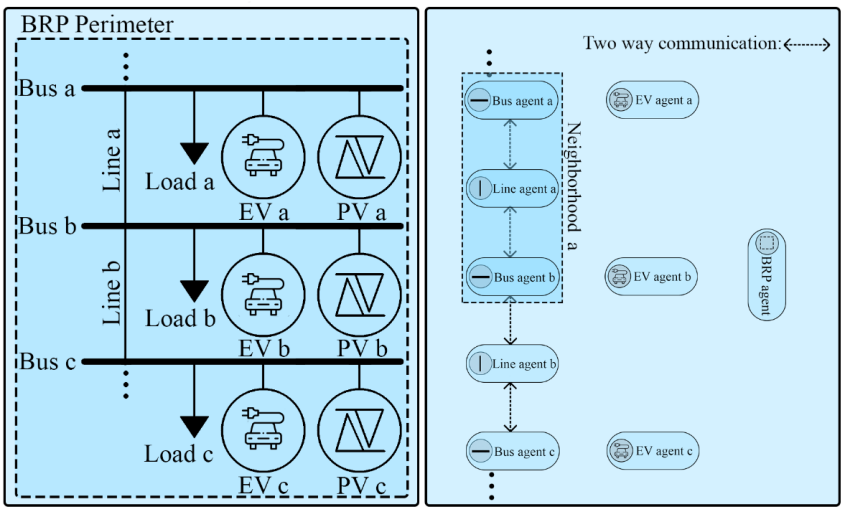

Figure 2.12: Section of a distribution network (left) and its agentified model (right) highlighting the neighborhood of line agent *a*.

export of power to the grid; or **null**: no electrical current congestion. A line agent also receives a set of requests $\mathcal{R}(t)$ from its neighboring agents at instant t.

During the *decision stage*, a line agent decides if it would forward its own criticality request or a request from the perceived set $\mathcal{R}(t)$ to its neighboring agents. If a line agent has a higher criticality than the criticalities of all the received neighboring requests, then the criticality to be forwarded Cr_f is the agent's own criticality $Cr(t)$, and issue to be forwarded $\mathcal{I}_f$ is set to agent's own issue $\mathcal{I}(t)$. However, if any of the received requests have a higher criticality than the line agent's own criticality (i.e., $\max_{Cr} \mathcal{R}(t) > Cr(t)$), then criticality of that request $\max_{Cr} \mathcal{R}(t)$ will be set as the criticality to be forwarded Cr_f, and the issue corresponding to the received request with highest criticality $\arg\max_{Cr} \mathcal{R}(t)$ will be set as the issue to be forwarded $\mathcal{I}_f$. Finally, the selected Cr_f and $\mathcal{I}_f$ values are forwarded to a line agent's neighbor in the *action* stage, only if Cr_f is non-zero.

Algorithm 2.1 AMAS line agent's functionality

Require: Electrical line's rated current $I_{ab,max}$
Require: Electrical line's threshold current $I_{ab,th}$
Require: Memory-based line criticality tuning parameter k_l
▷ **Perception stage**
$I_{ab}(t)$:= Perceived instantaneous line current from the sensor
$Cr(t)$:= Line agent's instantaneous criticality
$\mathcal{I}(t)$:= Line agent's instantaneous issue
$\mathcal{R}(t)$:= Set of requests received by line agent from its neighboring agents
if (Line agent has not been congested) **then**
    $Cr(t)$ is calculated using Equation (2.18)
else
    $Cr(t)$ is calculated using Equation (2.19)
end if
$\mathcal{I}(t)$:= Line agent's instantaneous issue
▷ **Decision stage**
Cr_f := Criticality value to be forwarded
$\mathcal{I}_f$:= Issue to be forwarded
$Cr_f := 0$
$\mathcal{I}_f$:= null
if ($\max_{Cr} \mathcal{R}(t) <= Cr(t)$) **then**
    $Cr_f := Cr(t)$
    $\mathcal{I}_f := \mathcal{I}(t)$
else
    $Cr_f := \max_{Cr} \mathcal{R}(t)$
    $\mathcal{I}_f$:= Issue corresponding to $\arg\max_{Cr} \mathcal{R}(t)$ request
end if
▷ **Action stage**
if ($Cr_f \neq 0$) **then**
    Forward $(Cr_f, \mathcal{I}_f)$ to neighboring agents
end if

Bus agent

Each bus agent is designed to maintain the voltage at its corresponding bus within a desired range i.e., Equation (2.11). An electrical bus can suffer from either an under-voltage issue (i.e., $V_{a,min} < V_a(t)$), or an over-voltage issue (i.e., $V_a(t) < V_{a,max}$). The objective of a bus agent is to prevent these issues at its electrical bus. A bus agent can achieve this objective by keeping its criticality equal to zero. Bus criticality of a bus agent $Cr_{b,a}(t)$, associated with electrical bus a, at instant t is calculated using the following *bus agent's criticality model*:

Bus agent's criticality model

$$Cr_{b,a}(t) = \begin{cases} 0 & \text{if } V^{-}_{a,th} \leq V_a(t) \leq V^{+}_{a,th} \\ \frac{V_a(t) - V^{-}_{a,th}}{V_{a,min} - V^{-}_{a,th}} & \text{if } V_{a,min} \leq V_a(t) < V^{-}_{a,th} \\ \frac{V_a(t) - V^{+}_{a,th}}{V_{a,max} - V^{+}_{a,th}} & \text{if } V_{a,max} \geq V_a(t) > V^{+}_{a,th} \\ 1 & \text{if } V_a(t) < V_{a,min} < V^{-}_{a,th} \\ 1 & \text{if } V_a(t) > V_{a,max} > V^{+}_{a,th} \end{cases} \tag{2.20}$$

In Equation (2.20), bus criticality is zero if the instantaneous rms voltage at bus a is in the range $[V^{-}_{a,th}, V^{+}_{a,th}]$. If this condition does not hold then either an over-voltage issue or an under-voltage issue is present at the electrical bus. It should be noted that here $V_{a,min} < V^{-}_{a,th} < V^{+}_{a,th} < V_{a,max}$. Terms $V^{-}_{a,th}$ and $V^{+}_{a,th}$ are negative and positive voltage thresholds respectively. If the instantaneous bus voltage is below $V^{-}_{a,th}$ or above $V^{+}_{a,th}$, the bus criticality starts increasing linearly. Bus criticality is maximum (i.e., $= 1$) when either over-voltage ($V_{a,max} < V_a(t)$) or under-voltage ($V_a(t) < V_{a,min}$) occurs. This is also apparent in Figure 2.13. A critical bus agent (i.e., $Cr_{b,a}(t) = 1$) can request EVs in its neighborhood for cooperation to reduce its criticality value.

Similar to line agents, the design of each bus agent also includes a feedback loop (i.e., criticality dictates the instantaneous action $\longrightarrow$ action impacts the instantaneous bus voltage $\longrightarrow$ bus voltage determines the next instant's bus criticality). This feedback loop can cause high oscillations in the system, which are highly undesirable. Thus, similar to line agents, the design of a bus agent also consists of a memory-based criticality model. A bus agent utilizes this memory-based criticality model to calculate its instantaneous memory-based criticality $Cr_{m,b,a}(t)$, once a bus has faced an over-voltage or an under-voltage issue. The *memory-based bus agent's criticality model* is given as follows:

Memory-based bus agent's criticality model

$$Cr_{m,b,a}(t) = k_b Cr_{b,a}(t) + (1 - k_b) Cr_{m,b,a}(t-1) \tag{2.21}$$

In Equation (2.21), the term k_b is the tuning parameter of the memory-based bus criticality model. The memory-based bus criticality $Cr_{m,b,a}(t)$ of a bus agent will con-

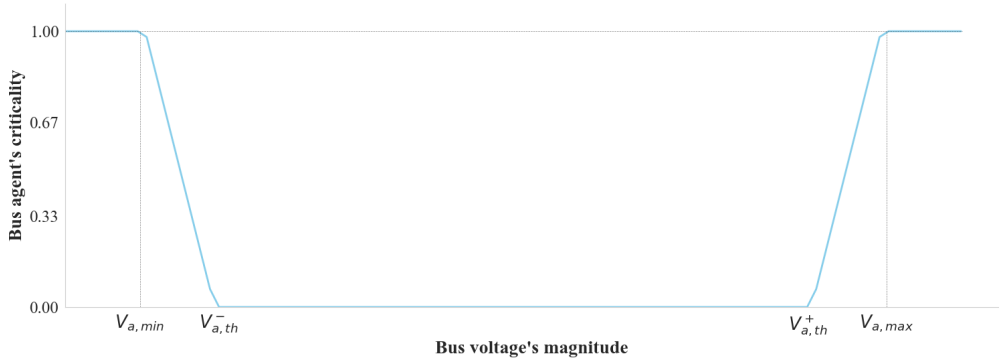

Figure 2.13: Relationship between bus criticality and electrical bus voltage's magnitude.

verge to its simple bus criticality value $Cr_{b,a}(t)$, if $Cr_{b,a}(t)$ is stationary. The rate of convergence depends on the selected value of k_b. A large value of k_b would result in faster convergence but might give rise to oscillations in the system. A smaller value of k_b would ensure none or minimal oscillations but the rate of convergence will be slower.

Just like line agents, a bus agent can communicate only in its neighborhood. A bus agent may include line agents and EV agents (corresponding to electrical lines and EVs connected to that bus in a physical distribution network). A visual representation of the neighborhood of a bus agent is shown in Figure 2.14. In Figure 2.14, neighborhood c is the neighborhood of bus agent c. Thus, bus agent c can only communicate with line agent b and EV agent c in Figure 2.14. Bus agent c will communicate the request with higher criticality between its own request and the request received from line agent b to EV agent c in Figure 2.14. Identical to line agents, the communication is happening in a single direction (either upstream or downstream). This means that the request received by bus agent c from line agent b may only be forwarded to EV agent c, and will never be sent back to line agent b. It should be noted that there is only one-way

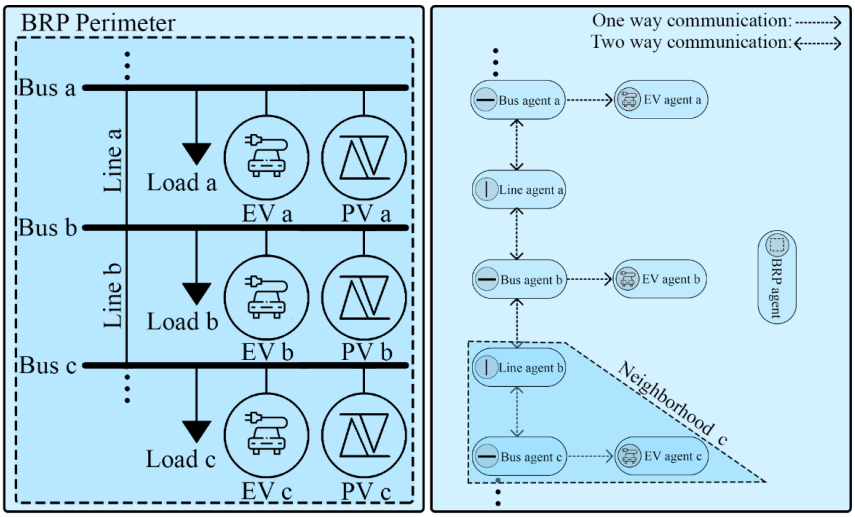

Figure 2.14: Section of a distribution network (left) and its agentified model (right) highlighting the neighborhood of bus agent a.

communication between a bus agent and an EV agent. This is because, in the designed system, the requests of line and bus agents are transferred to EV agents. Each EV agent utilizes the received information to make its instantaneous charging/discharging decision (i.e., the decision variable of the studied optimization problem). Thus, an EV agent does not require communicating any of its information to a bus agent in the designed adaptive multi-agent system.

Algorithm 2.2 AMAS bus agent's functionality

Require: Electrical bus' allowed minimum and maximum voltages $V_{a,min}$, $V_{a,max}$
Require: Electrical bus' threshold voltages $V^{-}_{a,th}$, $V^{+}_{a,th}$
Require: Memory-based bus criticality tuning parameter k_b

▷ **Perception stage**

$V_a(t) :=$ Perceived instantaneous bus voltage from the sensor
$Cr(t) :=$ Bus agent's instantaneous criticality
$\mathcal{I}(t) :=$ Bus agent's instantaneous issue
$\mathcal{R}(t) :=$ Set of requests received by bus agent from its neighboring agents
if (Bus agent has not been congested) **then**
    $Cr(t)$ is calculated using Equation (2.20)
    $\mathcal{I}(t) :=$ Bus agent's instantaneous issue
else
    $Cr(t)$ is calculated using Equation (2.21)
    $\mathcal{I}(t) :=$ Bus agent's instantaneous issue
end if

▷ **Decision stage**

$Cr_f :=$ Criticality value to be forwarded
$\mathcal{I}_f :=$ Issue to be forwarded
$Cr_f := 0$
$\mathcal{I}_f :=$ null
if ($\max_{Cr} \mathcal{R}(t) <= Cr(t)$) **then**
    $Cr_f := Cr(t)$
    $\mathcal{I}_f := \mathcal{I}(t)$
else
    $Cr_f := \max_{Cr} \mathcal{R}(t)$
    $\mathcal{I}_f :=$ Issue corresponding to $\arg\max_{Cr} \mathcal{R}(t)$ request
end if

▷ **Action stage**

if ($Cr_f \neq 0$) **then**
    Forward $(Cr_f, \mathcal{I}_f)$ to neighboring agents
end if

The communication message format followed by a bus agent is the same as a line agent. It communicates an ordered pair. Elements of this ordered pair are:

- **Criticality:** The first term of a communicated request (ordered pair) is a criticality value that ranges between 0 and 1.
- **Issue:** The second term in a communicated request (ordered pair) is an issue

status. This issue value can be either *over-voltage* or *under-voltage*. There are two possible issue values because the desired action from EVs in case of an under-voltage issue is to decrease their charging power. Whereas, when an over-voltage challenge is faced, EVs would be expected to increase their charging power. Thus, this issue helps determine how an EV agent can cooperate with the bus agent to reduce the criticality of the latter.

A bus agent, similar to line agents, should also cooperate with its neighboring system in an adaptive multi-agent system. For example, in Figure 2.14, if bus agent c is critical and it has also received a request from critical line agent b, then it has to decide which critical request it should forward to EV agent c. A bus agent can make this decision using the *comparison of criticalities* principle i.e., the bus agent will forward the request with the highest criticality to EV agent c in Figure 2.14. The algorithmic functionality of a bus agent is explained in Algorithm 2.2.

During the *perception* stage, a bus agent perceives instantaneous voltage magnitude at its bus. Based on this perceived voltage value, instantaneous criticality $Cr(t)$ is calculated by a bus agent in Algorithm 2.2. Term $I(t)$ in Algorithm 2.2 indicates the current issue a bus agent may face. This term can be either **over-voltage**: if the over-voltage issue is present at the bus; **under-voltage**: if the under-voltage issue is present at the bus; or **null**: if no issue is present at the bus. In Algorithm 2.2, a bus agent may also receive a set of requests $\mathcal{R}(t)$ from its neighboring agents at any given instant t. In the *decision* stage, a bus agent compares its criticality with the criticalities present in the received set of requests $\mathcal{R}(t)$. Functions $\max_{Cr} \mathcal{R}(t)$ and $\arg\max_{Cr} \mathcal{R}(t)$ return the maximum criticality and the request holding the maximum criticality present in the received set of requests $\mathcal{R}(t)$. Terms Cr_f and $\mathcal{I}_f$ stand for the criticality to be forwarded and the issue to be forwarded by a bus agent to its neighboring agents. In the *action* stage, the ordered pair (Cr_f, I_f) is forwarded to neighboring agents when Cr_f is non-zero.

Balance responsible party agent

The objective of a balance responsible party (BRP) agent is to minimize the energy mismatch during each of its imbalance settlement periods i.e., to minimize Equation (2.1). The instantaneous BRP criticality $Cr_{brp,a}(t)$ of BRP a is calculated using the given as follows *BRP agent's criticality model*:

BRP agent's criticality model

$$Cr_{brp,a}(t) = \max\left(\frac{\left|\left(\tilde{P}(N) - \frac{\sum_{j=1}^{t} P_{BRP}(j)}{t}\right)\Delta t\right|}{n' k_{brp}}, 1\right) \tag{2.22}$$

In Equation (2.22), term $n' \in (0, n]$ indicates the amount of time left before the end of the current BRP imbalance settlement period, and k_{brp} is the tuning parameter of $Cr_{brp,a}(t)$. Here, the idea is that a BRP will calculate the average consumption in its balance perimeter $\frac{\sum_{j=1}^{t} P_{BRP}(j)}{t}$, from the start of the present imbalance settlement period

till the current instant t. The BRP instantaneous criticality will depend on the difference between $\tilde{P}(N)$ (i.e., planned average consumption/production) and $\frac{\sum_{j=1}^{t} P_{BRP}(j)}{t}$ (i.e., average consumption/production till instant t). Evidently, the absolute value of this difference term (the numerator in Equation (2.22)) is directly proportional to a BRP agent's instantaneous criticality. Furthermore, $Cr_{brp,a}(t)$ is inversely proportional to n'. The relationship of instantaneous BRP criticality with n' and $\frac{\sum_{j=1}^{t} P_{BRP}(j)}{t}$ is shown in Figure 2.15.

In Figure 2.15, it can be seen that the instantaneous BRP criticality is zero when $\frac{\sum_{j=1}^{t} P_{BRP}(j)}{t} = \tilde{P}(N)$ i.e., when average production/consumption during the BRP imbalance settlement period is equal to the planned average production/consumption value during the BRP imbalance settlement period $\tilde{P}(N)$. The BRP criticality is non-zero when $\frac{\sum_{j=1}^{t} P_{BRP}(j)}{t} \neq \tilde{P}(N)$. Furthermore, when $n' = n$ (i.e., the imbalance settlement period has just started) then the BRP agent's criticality is closer to zero. However, if $n' \longrightarrow 0$ and $\frac{\sum_{j=1}^{t} P_{BRP}(j)}{t} \neq \tilde{P}(N)$, then the BRP criticality is near its maximum value. In case of a non-zero instantaneous criticality value, a BRP agent can request EV agents present in its neighborhood for cooperation.

The neighborhood of a BRP agent is shown in Figure 2.16. It can be seen in Figure 2.16 that the neighborhood of a BRP agent consists of all EV agents present inside the BRP perimeter. A BRP agent can communicate with these EV agents when required. There is a two-way communication link between a BRP and an EV agent. The communication message format (i.e., contents of the sent message) depends on the direction of the communication. An EV agent does not need to communicate its criticality to a BRP agent in the designed system (as an EV agent is the decision-making entity). However, each EV agent needs to communicate its instantaneous charging/discharging power to its BRP agent. A BRP agent uses the sum of instantaneous EVs' charging/discharging

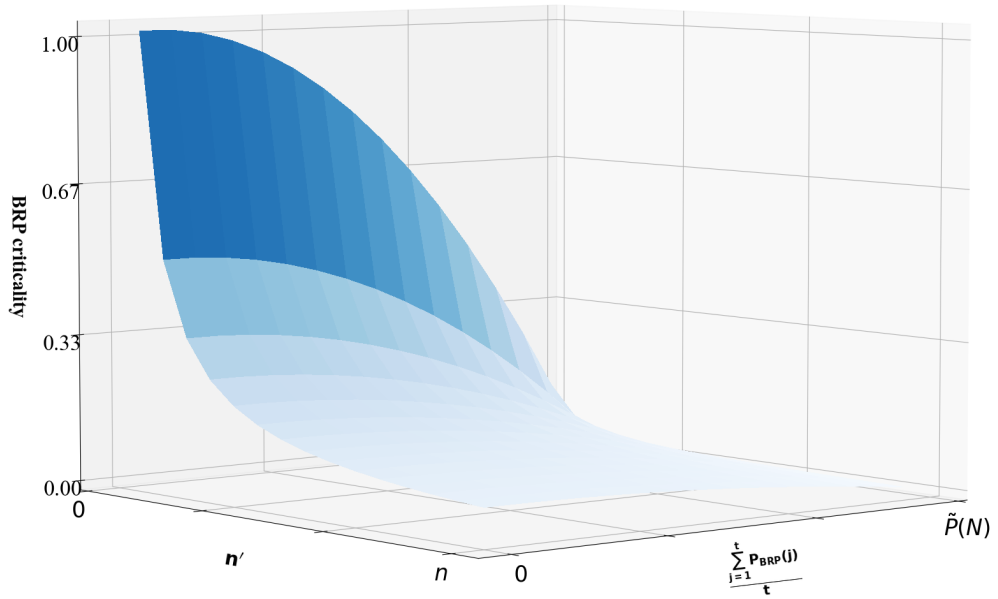

Figure 2.15: Relationship of BRP criticality with n' and $\frac{\sum_{j=1}^{t} P_{BRP}(j)}{t}$.

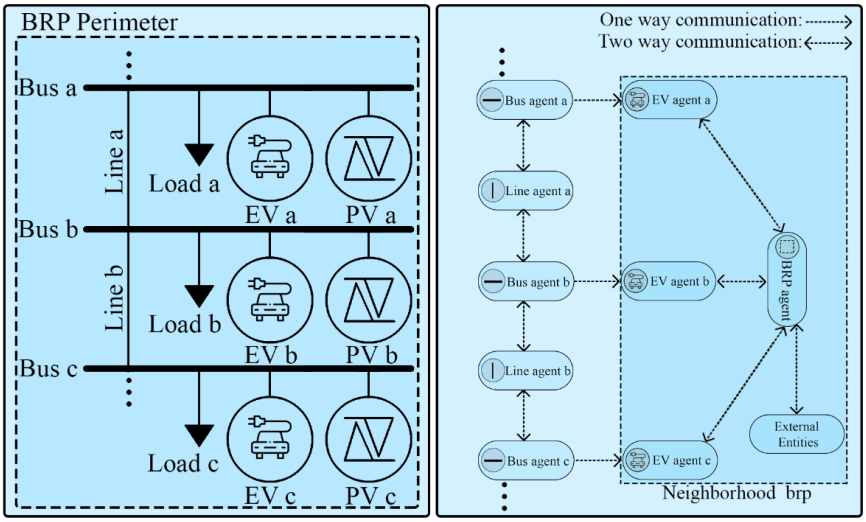

Figure 2.16: Section of a distribution network (left) and its agentified model (right) highlighting the neighborhood of BRP agent.

power to calculate instantaneous production/consumption in its perimeter. Therefore, the message sent by an EV agent to a BRP agent consists of a single value only i.e., the instantaneous EV charging/discharging power $P_{e,a}(t)$. On the other hand, the message (request) a BRP agent sends to an EV agent for cooperation consists of an ordered pair. The elements of this communication ordered pair are:

- **Criticality:** The first term is the instantaneous BRP criticality value between 0 and 1.
- **Issue:** The second term in the current issue faced by a BRP agent. This issue value can be either *over-consumption* (BRP perimeter consuming more than the planned average consumption value), or *under-consumption* (BRP perimeter consuming lower than the planned average consumption value). In case of over-consumption, EVs should decrease their charging powers. Whereas, in case of under-consumption, EVs should increase their charging powers.

A BRP agent also communicates with a number of external entities (i.e., entities that have not been modeled as agents in the designed AMAS). These external entities are as follows:

- **Households:** A BRP agent receives instantaneous consumption from households present in its balance perimeter. This received data is used to calculate the total instantaneous production/consumption.
- **Photovoltaics:** Photovoltaics present in a balance perimeter are producing electrical energy. They communicate their instantaneous energy production values to their corresponding BRP agent. A BRP agent uses the received information to calculate the total instantaneous production/consumption.

Algorithm 2.3 AMAS BRP agent's functionality

Require: Duration of an imbalance settlement period n
Require: BRP criticality tuning parameter k_{brp}
▷ **Perception stage**
1: $Cr(t) \coloneqq$ BRP agent's instantaneous criticality
2: $\mathcal{I}(t) \coloneqq$ BRP agent's instantaneous issue
3: Perceive $P_{load,a}(t)$ from each household a
4: Perceive $P_{PV,a}(t)$ from each PV a
5: Perceive $P_{e,a}(t)$ from each EV a
6: Calculate $Cr(t)$ using Equation (2.22)
7: Determine BRP agent's instantaneous issue $\mathcal{I}(t)$
▷ **Decision stage**
8: $Cr_f \coloneqq$ Criticality value to be forwarded
9: $Cr_f \coloneqq$ Cr(t)
10: $\mathcal{I}_f \coloneqq$ Issue to be forwarded
11: $\mathcal{I}_f \coloneqq \mathcal{I}(t)$
▷ **Action stage**
12: **if** ($Cr_f \neq 0$) **then**
13: Forward $(Cr_f, \mathcal{I}_f)$ to neighboring agents
14: **end if**

- **Transmission system operator:** A BRP agent also communicates with its transmission system operator (TSO) to share its day-ahead planned average production/consumption schedule $\tilde{P}(N)$. It may also communicate its encountered total commitment mismatch to TSO.

The functionality of a BRP agent is described in Algorithm 2.3. In Algorithm 2.3, the BRP agent observes the production/consumption of each household, PV and EV during the perception stage. Based on the observed data, instantaneous BRP criticality is calculated using Equation (2.22). The possible values of the instantaneous issue variable $\mathcal{I}(t)$ are: **over-consumption**: when real-time average production/consumption is greater than the planned value; **under-consumption**: when real-time average production/consumption is lower than the planned value; or **null**: when real-time average production/consumption is equal to the planned value. Variables Cr_f and $\mathcal{I}_f$ are put equal to $Cr(t)$ and $\mathcal{I}(t)$ respectively during the *decision* stage. Finally, if the criticality to be forwarded is non-zero, the cooperation request is sent to all neighboring (EVs) agents during the *action* stage.

Electric vehicle agent

The main objective of each EV agent is to ensure that the prosumer objective given in Equation (2.17) is satisfied. This would have been a straightforward objective if an EV agent would not have to cooperate with its neighborhood agents. In the studied system, an EV agent must ensure that the objective of its BRP agent along with the DSO constraints is satisfied. Thus, an EV agent also needs to utilize the *comparison of criticalities* principle to decide whether it should help itself or one of its neighboring

agents. For that, an EV agent is required to calculate its criticality first. An EV agent e, connected to electrical bus a, calculates its own criticality using the *EV agent's criticality model* given as follows:

EV agent's criticality model

$$Cr_{e,a}(t) = \begin{cases} \max\left(\frac{\left(SoC_{e,a,depart} - SoC_{e,a}(t)\right)E_{e,a,bat}}{k_e \left|t_{e,a,depart}\right| P_{e,a,max}}, 1\right) & \text{if } SoC_{e,a}(t) < SoC_{e,a,max} \\ 0 & \text{if } SoC_{e,a,max} \leq SoC_{e,a}(t) \end{cases} \tag{2.23}$$

Terms $SoC_{e,a,depart}$, $E_{e,a,bat}$, and $P_{e,a,max}$ represent the desired final state of charge, the battery capacity, and the maximum instantaneous charging power of EV e, which is connected to electrical bus a. The tuning parameter of an EV's criticality is represented by k_e here. The instantaneous criticality $Cr_{e,a}(t)$ of EV e, given in Equation (2.23), is linked to two time-varying variables i.e., $SoC_{e,a}(t)$ and $|t_{e,a,depart}|$. The relationship between these time-varying variables and EV's criticality is shown in Figure 2.17. Term $|t_{e,a,depart}|$ indicates the time remaining before the departure of EV e. This term is inversely proportional to an EV agent's instantaneous criticality $Cr_{e,a}(t)$ for a given value of EV's instantaneous state of charge $SoC_{e,a}(t)$. On the contrary, $SoC_{e,a}(t)$ is directly proportional to $Cr_{e,a}(t)$ for a fixed value of $|t_{e,a,depart}|$. It can be seen in Figure 2.17 that the criticality value is maximum (i.e., $Cr_{e,a}(t) \longrightarrow 1$) when $|t_{e,a,depart}| \longrightarrow 0$, and the criticality value is around minimum (i.e., $Cr_{e,a}(t) \longrightarrow 0$) if $|t_{e,a,depart}| \longrightarrow \infty$. Additionally, the criticality value is non-zero when $SoC_{e,a}(t) \longrightarrow SoC_{e,a,min}$, and the criticality value is around its minimum (i.e., $Cr_{e,a}(t) \longrightarrow 0$) if $SoC_{e,a}(t) \longrightarrow SoC_{e,a,depart}$.

The neighborhood definition of an EV agent is shown in Figure 2.18. It can be seen that an EV agent includes a bus agent and a BRP agent in its neighborhood. For

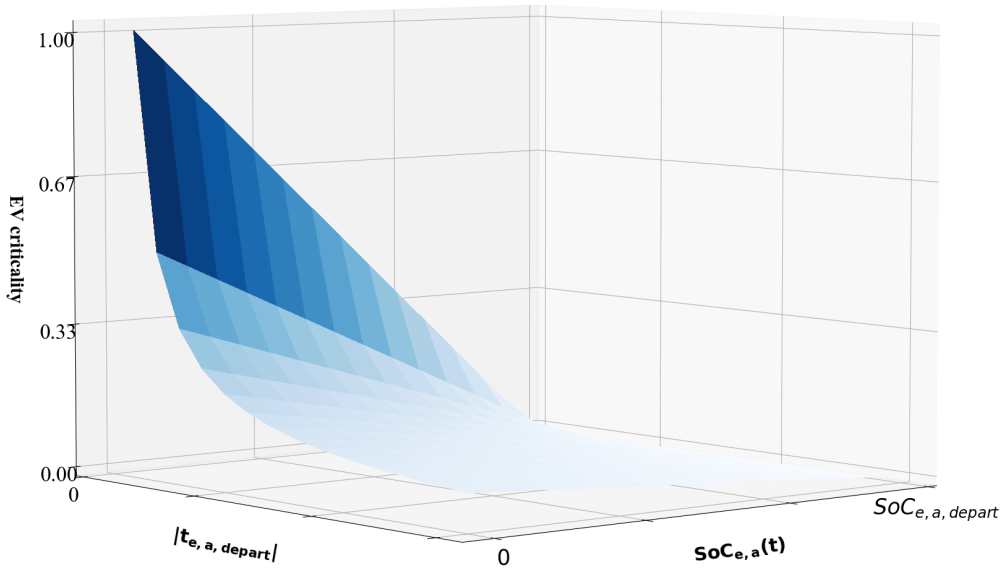

Figure 2.17: Relationship of EV criticality with $|t_{e,a,depart}|$ and $SoC_{e,a}(t)$.

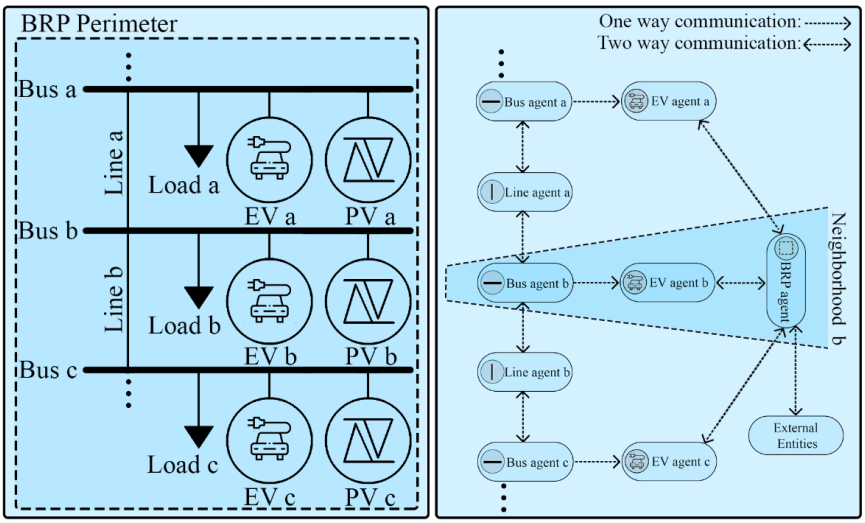

Figure 2.18: Section of a distribution network (left) and its agentified model (right) highlighting the neighborhood of EV agent.

example, in Figure 2.18, EV *b* is present in a BRP's perimeter, and it is connected to an electrical bus *b*. Thus, in the agentified model, EV agent *b* included *BRP* agent and bus agent *b* as its neighboring agents. There is a one-way communication link between an EV and its neighboring bus agent. The bus agent is sending (or forwarding) a cooperation request to the connected EV agent. A BRP agent is also communicating cooperation requests in case of need to all EV agents in its neighborhood. Each EV agent is also communicating with its BRP. Each EV communicates its instantaneous charging/discharging power to its BRP. Each EV agent is designed to cooperate with its neighboring agents. At any given instant, an EV agent may receive cooperation request(s) from its neighboring agents. Thus, at each instant, an EV agent must decide whether to satisfy its objective or help a neighboring agent. This decision is made by an EV agent using the *comparison of criticalities* principle. Thus, an EV agent changes its instantaneous charging/discharging power, with respect to its past instant's charging/discharging power, based on the magnitude of the highest instantaneous criticality and the communicated issue corresponding to this criticality. This *EV agent's power calculation model* is presented below:

EV agent's power calculation model

$$P_{e,a}(t) = P_{e,a}(t-1) + \lambda Cr_{max}(t) P_{e,a,max} \tag{2.24}$$

In Equation (2.24), $P_{e,a}(t)$ is the instantaneous charging power of EV e, $P_{e,a,max}$ is the maximum instantaneous charging power of EV e, and $Cr_{max}(t)$ is the value of highest criticality at instant t. It should be noted that the variable $P_{e,a}(t)$ is bounded between $P_{e,a,min}$ and $P_{e,a,max}$ i.e., $P_{e,a}(t) \in [P_{e,a,min}, P_{e,a,max}]$. As stated earlier, the direction of change in $P_{e,a}(t-1)$ depends on the issue associated with $Cr_{max}(t)$. The value of λ in Equation (2.24) determines if the EV agent will be increasing its instantaneous

Issue type	Issue associated with $Cr_{max}(t)$	λ
EV	Low state of charge	1
BRP	Over-consumption	-1
	Under-consumption	1
Bus	Under-voltage	-1
	Over-voltage	1
Line	High current (import)	-1
	High current (export)	1

Table 2.1: Possible issues associated with the highest criticality and their impact on the EV agent's power calculation model.

charging power or decreasing its instantaneous charging power compared to the previous instant. Here, $\lambda \in \{-1, 1\}$. If $\lambda = 1$, then the EV agent will be increasing its instantaneous charging power compared to the previous instant. If $\lambda = -1$, then the EV agent will be decreasing its instantaneous charging power compared to the previous instant. The possible values of λ in relation to the issue associated with the highest observed criticality by the EV agent are given in Table 2.1.

It is evident in Table 2.1 that an EV agent can face seven possible issues (divided into five groups based on the origin of these issues). If an EV agent has the highest criticality due to a low state of charge then $P_{e,a}(t)$ should be greater than $P_{e,a}(t-1)$. If a BRP agent is facing an over-consumption issue then $P_{e,a}(t)$ is desired to be lower than the $P_{e,a}(t-1)$. On the other hand, if there is under-consumption in the BRP perimeter then $P_{e,a}(t)$ must be greater than $P_{e,a}(t-1)$. The highest criticality can also be associated with electrical bus issues (i.e., over- or under-voltage) and electrical line issues (i.e., high import or export current). In case of under-voltage or high import current issues, an EV should decrease its charging power. On the other hand, an EV should increase its charging power when an over-voltage or high export current issue is faced.

In the studied optimization problem, it is possible that an antagonistic situation may arise. An *antagonistic situation* is defined as the situation in which two critical agents are demanding opposite cooperative actions from an EV agent. For example, it is possible that a BRP agent is facing an under-consumption issue. Thus, it would request EVs to charge more. But at the same time, a line agent could also face high import current issues and it would request EVs to decrease their charging powers. Therefore, it gives rise to a situation when EV agents are requested two opposite cooperative actions. To tackle this challenge, instead of using the model in Equation (2.24) for instantaneous power calculation, an EV agent utilizes the following model:

EV agent's power calculation model (considering antagonistic scenarios)

$$P_{e,a}(t) = P_{e,a}(t-1) \pm (Cr_{max}(t) - hCr_{ant}(t))\, P_{e,a,max} \tag{2.25}$$

In Equation (2.25), $P_{e,a}(t)$ depends on two new variables in comparison to Equation (2.24). These variables are the antagonistic request's instantaneous criticality $Cr_{ant}(t)$, and the *h*-value. Term *h*-value is modeled as a function of the instantaneous memory-based line criticality $Cr_{m,l,ab}(t)$. This dependency is made to prioritize electrical line congestion in case an EV agent has received antagonistic requests. Indeed, it is assumed that the distribution network's stability is more critical than optimizing a BRP's cost. Furthermore, it is also assumed that the global stability ensured by a BRP remains guaranteed even when EVs are giving preference to solving a congestion issue over the optimization of a BRP's cost. That is why when a BRP and a line agent request opposite cooperative actions from an EV, priority will be given to the line agent's request. The instantaneous *h*-value can be calculated using the following model:

h-value model

$$h(Cr_{m,l,ab}(t)) = \begin{cases} 1 + \frac{(\alpha-1)Cr_{m,l,ab}(t)}{\gamma_{l,min}} & \text{if } Cr_{m,l,ab}(t) < \gamma_{l,min} < \gamma_{l,max} \\ e^{\frac{Cr_{m,l,ab}(t)}{pa+p_b}} & \text{if } \gamma_{l,min} \leq Cr_{m,l,ab}(t) \leq \gamma_{l,max} \\ 1 - \frac{\beta Cr_{m,l,ab}(t)}{1-\gamma_{l,max}} & \text{if } \gamma_{l,min} < \gamma_{l,max} < Cr_{m,l,ab}(t) \end{cases} \tag{2.26}$$

The values of p_a and p_b, in Equation (2.26), can be calculated as follows:

$$p_a = \frac{\gamma_{l,max}}{ln(\beta) - ln(\alpha)} \tag{2.27}$$

$$p_b = \frac{ln(\beta\gamma_{l,min}) - ln(\alpha\gamma_{l,max})}{\gamma_{l,max} - \gamma_{l,min}} \tag{2.28}$$

Also, $\gamma_{l,min}$ and $\gamma_{l,max}$ are minimum and maximum thresholds. When $Cr_{m,l,ab}(t)$ is between these two threshold values, *h*-value behaves exponentially. Whereas, *h*-value behaves linearly otherwise i.e., when $Cr_{m,l,ab}(t) \notin [\gamma_{l,min}, \gamma_{l,max}]$. Terms α and β are scaling parameters that range between 0 and 1. The *h*-value also ranges between 0 and 1. The relationship between *h*-value and $Cr_{m,l,ab}(t)$ is also visually represented in Figure 2.19. This *h*-value basically decides the level of priority that should be given to an antagonistic request when an electrical line is congested in the network. For higher values of $Cr_{m,l,ab}(t)$, it can be seen in 2.19 that the *h*-value is decreasing exponentially. Thus, an EV agent would be giving more priority to the request of a line agent when deciding its next instant's charging/discharging power.

The detailed algorithmic functionality of an EV agent is presented in Algorithm 2.4. During the *perception* stage, an EV agent observes the instantaneous state of charge of its EV $SoC_{e,a}(t)$. Based on the observed $SoC_{e,a}(t)$ value, an EV agent calculates its instantaneous criticality. The agent also receives cooperation requests from

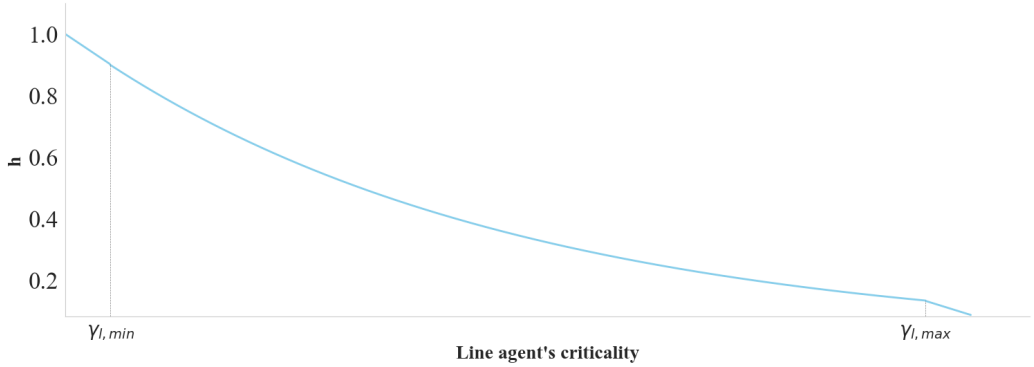

Figure 2.19: Relationship between a line agent's memory-based criticality and its h-value.

its neighborhood. Moving forward, during the *decision* stage, the agent finds the criticality of the most critical agent in its neighborhood (including itself). It also finds the highest criticality of the received antagonistic requests (if any). Finally, the agent calculates the instantaneous charging/discharging power of its corresponding EV using Equation (2.25). This power is set as the EV's instantaneous charging/discharging power and communicated to the neighboring BRP during the *action* stage.

This concludes the design of the proposed adaptive multi-agent system to handle the real-time grid balancing optimization problem in smart grids. The detailed func-

Algorithm 2.4 AMAS EV agent's functionality

Require: Desired SoC at departure time $SoC_{e,a,depart}$
Require: Battery capacity $E_{e,a,bat}$
Require: Minimum charging power $P_{e,a,min}$
Require: Maximum charging power $P_{e,a,max}$
Require: h-value scaling parameters α and β
▷ **Perception stage**
$SoC_{e,a}(t)$:= Perceived instantaneous EV's state of charge
$Cr(t)$:= EV agent's instantaneous criticality
$\mathcal{I}(t)$:= EV agent's instantaneous issue
$\mathcal{R}(t)$:= Set of requests received by EV agent from its neighboring agents
Calculate $Cr(t)$ using Equation (2.23)
Determine EV agent's instantaneous issue $\mathcal{I}(t)$
▷ **Decision stage**
$P_{e,a}(t)$:= EV's instantaneous charging power
Cr_{max} := Find the highest criticality among $Cr(t)$ and criticalities in $\mathcal{R}(t)$
$\mathcal{I}(t)$:= Issue associated with Cr_{max}
Cr_{ant} := Find the antagonistic criticality among $Cr(t)$ and criticalities in $\mathcal{R}(t)$
Calculate the instantaneous h-value using Equation (2.26)
$P_{e,a}(t)$:= Calculate EV's instantaneous charging power using Equation (2.25)
▷ **Action stage**
Set EV's instantaneous charging power equal to $P_{e,a}(t)$
Communicate $P_{e,a}(t)$ to the neighboring BRP agent

tionalities of all modeled agent types (i.e., line, bus, BRP, and EV) have been presented in the section. The proposed AMAS system has been called decentralized because each network entity that encounters an issue (e.g., a node with under-voltage, an electrical line with congestion, or a BRP with commitment mismatch) deals with this issue by sending messages to the flexible entities (EVs here), and each EV adapts its charging strategy accordingly. This approach differs from the centralized control implemented by a DSO, where the system identifies grid issues and determines how to respond. The presented system is also scalable, model-free, real-time, and generic (adaptable to other smart grid applications). The upcoming chapter presents a thorough evaluation of the proposed system through comparative deterministic and pseudo-stochastic studies. Performance comparison will be conducted with two baseline EV charging optimization strategies, which are also discussed in the subsequent chapter.

2.5 Conclusion

This chapter presents a novel adaptive multi-agent system to tackle the studied smart grid optimization problem. In the studied optimization problem, the objective was utilizing the flexible entities (i.e., electric vehicles) to provide ancillary services to the BRP by minimizing its commitment mismatch error during each imbalance settlement period. Along with performing optimization, the aim was also to satisfy the constraints of different market actors such as DSOs and prosumers. The initial part of this chapter focused on presenting the detailed mathematical formulation of the problem under study, along with a review of the related literature and a clear description of the problem's scope within this thesis. Subsequently, a comprehensive discussion on the theory of adaptive multi-agent systems was provided, highlighting their distinctions from simple multi-agent systems and emphasizing how self-organization in adaptive multi-agent systems can enhance the system's flexibility. The philosophy of adaptive multi-agent systems was then applied to design a fully decentralized control system for addressing the studied smart grid optimization problem. The developed system comprises distinct agent types, each with its own objective within the system, and their functionalities were elaborated upon. In the next chapter, a detailed evaluation of the proposed system will be conducted, involving comparisons with other baseline strategies through simulation-based experiments. This evaluation aims to provide insights into the performance and standing of the proposed system as well as to identify potential areas for further improvement.

Chapter 3

Evaluating adaptive multi-agent system for grid balancing

It is better to change an opinion than to persist in a wrong one.

Socrates

Summary

The primary focus of this chapter is to conduct a comprehensive evaluation of the adaptive multi-agent system proposed in the preceding chapter, specifically tailored to manage real-time grid balancing operations. This evaluation is carried out through extensive deterministic and pseudo-stochastic simulation-based experiments. The performance of the proposed adaptive multi-agent system is compared with two baseline electric vehicle charging optimization strategies, which are detailed in this chapter. The comparison comprises factors such as optimality, satisfaction of constraints, and scalability of these approaches. Pseudo-stochastic simulation case studies are conducted to assess the system's performance under uncertain conditions. The objective of this chapter is also to identify potential areas for further improving the system's performance, if necessary.

Contents

The objective of this chapter is to assess the performance of the proposed adaptive multi-agent system introduced in Chapter 2. To achieve this, two baseline optimization strategies are presented in Section 3.1. These strategies consist of the uncontrolled approach (where electric vehicle charging is not controlled by any entity) and the centralized mixed-integer linear programming optimization strategy (where a centralized operator determines the charging strategy for each electric vehicle using mixed-integer linear programming optimization). At first, deterministic simulation-based experiments are conducted to compare the proposed system's performance with these baseline strategies in Section 3.2. The simulation case study settings and results are discussed in detail in the mentioned section. Additionally, the impact of real-life uncertainties on the system is investigated through pseudo-stochastic simulation-based experiments in Section 3.3. This section analyzes how the system performs under uncertain conditions. To conclude this chapter, a summary of the observations made is provided, and potential avenues for further system enhancements are highlighted in Section 3.4.

3.1 Baseline optimization strategies

In this section, alternative control strategies are presented. It is essential to evaluate the performance of any newly developed decentralized system. Thus, the suggested strategies in this section will be used to benchmark the proposed adaptive multi-agent system. The comparison will help to identify the advantages and drawbacks of the proposed multi-agent system. It will also help determine how the designed adaptive multi-agent system can be improved. Two strategies are selected as baselines:

- Uncontrolled strategy
- Centralized MILP optimization strategy

Uncontrolled strategy:

As the name suggests, the instantaneous charging power of each EV $P_{e,a}(t)$ is not controlled in the *uncontrolled strategy* [95]. In this charging strategy, each EV e starts charging at its rated power $P_{e,a,max}$ as soon as it is connected to the grid. This strategy may lead to instability in the network due to peak load demand. Furthermore, a BRP cannot exploit the existing EVs in its perimeter for its benefit, as the instantaneous charging powers of EVs can not be controlled. There is no vehicle-to-grid (V2G) present in this scenario as well. Thus, if a BRP faces an under-production issue (due to lower-than-expected PV production or higher-than-expected loads' consumption) during one of its imbalance settlement periods, it can never minimize this under-production mismatch issue by requesting idle EVs to discharge. It means that strategy would lead to very sub-optimal solutions. These solutions will be considered as upper bounds while evaluating the performance of the proposed adaptive multi-agent system. It should be noted that although this strategy may be the most problematic strategy grid-wise, it is the simplest one for prosumers (as no controller is required to

control each EV's charging/discharging power). This strategy represents a business-as-usual scenario (a scenario in which future smart grids would be operating without any control of an increasing number of EVs). Thus, the solutions obtained through this strategy can represent a good upper bound to evaluate the performance of other control optimization strategies. The functionality of each EV, when it follows this uncontrolled strategy, is presented in Algorithm 3.1.

Algorithm 3.1 Uncontrolled strategy (each EV)

Require: Desired SoC at departure time $SoC_{e,a,depart}$
Require: Rated charging power of EV $P_{e,a,max}$
$SoC_{e,a}(t) \coloneqq$ SoC at instant t
$t_{arrive} \coloneqq$ EV's arrival time
$t_{depart} \coloneqq$ EV's departure time
for $t = 1, 2, 3, ..., T$ **do**
    if $(t \geq t_{arrive}$ & $t \leq t_{depart})$ & $(SoC_{e,a}(t) < SoC_{e,depart})$ **then**
        Charge at $P_{e,a,max}$
    else
        Do not charge
    end if
end for

In Algorithm 3.1, each EV starts charging at its rated power $P_{e,a,max}$ as long as it is connected to the grid (i.e., $t \geq t_{arrive}$ & $t \leq t_{depart}$) and it has yet to achieve its desired state of charge at its departure time (i.e., $SoC_{e,a}(t) < SoC_{e,depart}$)). As soon as the EV e either departs (i.e., $t \geq t_{depart}$) or it achieves its desired state of charge (i.e., $SoC_{e,a}(t) = SoC_{e,depart}$), it will stop charging.

Centralized MILP optimization strategy:

Mixed-integer linear programming (MILP) belongs to the class of mathematical optimization problems. It has found several practical applications, such as production planning [184], demand response optimization [39], unit commitment [30] etc. The use of MILP to optimize power flows in an electrical grid has been suggested in the literature for several decades. A mixed integer linear programming (MILP) optimization problem is of the form [23]:

Standard MILP formulation

$$\min c^T x \tag{3.1}$$

$$\mathcal{A}x = b \quad \text{s.t.} \quad x \in \mathbf{Z}^+ \tag{3.2}$$

In the above-given formulation, if all variables x are required to be integers then it becomes pure integer linear programming (ILP). On the other hand, if $x \in \{0, 1\}$ then it becomes 0-1 linear programming. The original smart grid optimization problem,

presented in Section 2.1, belongs to the quadratic constrained programming (QCP) class of optimization problems [4]. Equation (2.9) involves a quadratic term (i.e., product of voltages). The studied objective function in Equation (2.1) also involves an absolute term. Furthermore, the decision variable $P_{e,a}(t) \in [P_{e,a,min}, P_{e,a,max}]$ can be negative (when an EV is discharging). Thus, the original problem in 2.1 needs to be adapted to apply mixed-integer linear optimization.

Objective function linearization

As explained earlier, the original objective function stated in Equation (2.1) is not linear due to the absolute operator. To apply MILP optimization, this objective function must be linearized first. The absolute term in Equation (2.1) can be linearized by setting it equal to variable Q, and then constraining this assumed variable. This is done as follows:

Smart charging problem's linearized objective function

Objective function:

$$\min E_{mis} = \min \sum_{N=1}^{N_{end}} \left| \left(\tilde{P}(N) - \frac{\sum_{t=1}^{n} P_{BRP}(t)}{n} \right) \Delta t \right| \tag{3.3}$$

is equivalent to:

$$\min E_{mis} = \min \sum_{N=1}^{N_{end}} Q \Delta t \tag{3.4}$$

subject to the constraints:

$$\left(\tilde{P}(N) - \frac{\sum_{t=1}^{n} P_{BRP}(t)}{n} \right) \leq Q \tag{3.5}$$

$$\left(\frac{\sum_{t=1}^{n} P_{BRP}(t)}{n} - \tilde{P}(N) \right) \leq Q \tag{3.6}$$

Decision variable linearization

As vehicle-to-grid (V2G) is considered in the studied optimization problem, the decision variable $P_{e,a}(t)$ of the studied optimization problem can hold a negative value in case an EV is discharging. However, this decision variable should be a positive integer, as specified in Equation (3.2). This problem is solved by dividing our decision variable $P_{e,a}(t)$ into two parts: $P_{e,a,chrg}(t)$, and $P_{e,a,dischrg}(t)$. The instantaneous power $P_{e,a}(t)$ of EV e is the sum of its instantaneous charging power $P_{e,a,chrg}(t)$ and the negative of its instantaneous discharging power $P_{e,a,dischrg}(t)$. The instantaneous charging power $P_{e,a,chrg}(t)$ of EV e ranges between 0 and its rated charging power $P_{e,a,max}$.

Whereas, the instantaneous discharging power $P_{e,a,dischrg}(t)$ of EV e varies between 0 and its rated discharging power $P_{e,a,min}$. Both $P_{e,a,max}$ and $P_{e,a,dischrg}(t)$ are positive integers. To make sure that an EV does not charge and discharge simultaneously, a binary decision variable $\rho \in \{0,1\}$ is introduced. The mathematical formulation of this decision variable linearization is given in Equation (3.7) as follows:

Decision variable linearization

$$\begin{aligned} P_{e,a}(t) = \rho P_{e,a,chrg}(t) - (1-\rho) P_{e,a,dischrg}(t) \quad \text{s.t.} \quad \rho \in \{0,1\}, \\ P_{e,a,chrg}(t) \in [0, P_{e,a,max}], \\ P_{e,a,dischrg}(t) \in [0, P_{e,a,min}] \end{aligned} \tag{3.7}$$

Constraint linearization

Finally, one must also linearize Equation (2.9). The right hand side of Equation (2.9) can also be written as:

$$S_{ab}(t) = P_{a,b}(t) + iQ_{a,b}(t) = V_a(t) I^*_{ab}(t) = V_a(t) \left(Y^*_{ab} V^*_b(t)\right) \tag{3.8}$$

where $S_{ab}(t)$ is the instantaneous real power flowing from bus a to bus b. The voltage can be represented in polar coordinates as the product of magnitude and complex exponential i.e. $|V|\, e^{i(\omega t+\psi)}$, where ω represents the voltage angular frequency and ψ is the voltage angle. Furthermore, $Y_{ab} = G_{ab} + iB_{ab}$, where G_{ab} is the conductance of the electrical line connecting bus a and bus b, and B_{ab} is the susceptance of the electrical line between bus a and bus b.

$$\begin{aligned} S_{ab}(t) &= \left(|V_a| e^{i(\omega t+\psi_a)}\right) \sum_b \left((G_{ab} + iB_{ab})^* \left(|V_b| e^{i(\omega t+\psi_b)}\right)^* \right) \\ &= \sum_b \left(|V_a||V_b| e^{i(\omega t+\psi_a)} e^{-i(\omega t+\psi_b)}\right) (G_{ab} - iB_{ab}) \\ &= \sum_b \left(|V_a||V_b| e^{i(\psi_a-\psi_b)}\right) (G_{ab} - iB_{ab}) \\ &= \sum_b \left(|V_a||V_b| \left(cos(\psi_a - \psi_b) + isin(\psi_a - \psi_b)\right)\right) (G_{ab} - iB_{ab}) \end{aligned} \tag{3.9}$$

Comparing Equation (3.8) with Equation (3.9), active and reactive power flows can be written as follows:

$$P_{ab}(t) = G_{ab}|V_a|^2 - |V_a||V_b| \left(G_{ab}\, cos(\psi_a - \psi_b) - B_{ab}\, sin(\psi_a - \psi_b)\right) \tag{3.10}$$

$$Q_{ab}(t) = B_{ab}|V_a|^2 - |V_a||V_b| \left(G_{ab}\, sin(\psi_a - \psi_b) + B_{ab}\, cos(\psi_a - \psi_b)\right) \tag{3.11}$$

Finally, to obtain fully linearized active and reactive power flow equations, the following assumptions are made:

- **Small angle approximation**: It is assumed that voltage difference values are small enough to occupy a linear region of the sine function, i.e. $(\psi_a - \psi_b) \approx sin(\psi_a - \psi_b)$
- **Unit voltage magnitude**: It is also assumed that the voltage magnitude $|V_a|$ is sufficiently close to one per unit value, i.e. $|V_a| \approx 1$.

After applying these assumptions, the following two separate linear power flow equations are obtained:

Linearized power flow constraints

$$P_{ab}(t) = G_{ab}(t)\,(V_a(t) - V_b(t)) + B_{ab}(t)\,(\psi_a(t) - \psi_b(t)) \tag{3.12}$$

$$Q_{ab}(t) = B_{ab}(t)\,(V_a(t) - V_b(t)) + G_{ab}(t)\,(\psi_b(t) - \psi_a(t)) \tag{3.13}$$

Based on the earlier described constraints in Section 2.1 combined with linearized constraints and objective function explained in this section, a feasible set (Feasible set 3.1) can be obtained. This feasible set can be solved as a mixed integer linear programming (MILP) optimization problem. As this MILP formulation comes under the category of centralized optimization, the solution obtained through it will be considered as the lower bound to evaluate the performance of the proposed adaptive multi-agent system.

Feasible set 3.1: Linearized smart charging formulation

Objective function in Equation (3.4)

DSOs constraints in Equations (2.10)−(2.13)

Prosumers constraints in Equations (2.14)−(2.17)

Decision variable constraints in Equation (3.7)

Objective function linearization constraints in Equations (3.5)−(3.6)

Network's physical constraints in Equations (2.5)−(2.8) & (3.12)-(3.13)

3.2 Deterministic simulation-based experimentation

In this section, an evaluation of the proposed adaptive multi-agent system to optimize energy flows in the smart grid under deterministic conditions is done. In the beginning, simulation-based experimentation settings are discussed. Afterward, a comparison of the results obtained through the adaptive multi-agent system presented in Section 2.4 is made with the baselines presented in Section 3.1.

Simulation-based experimentation settings

A careful design of the experiment is required first to draw a comparison among different smart grid optimization strategies. It includes modeling the studied distribution network, selecting study parameters, preparing datasets, implementing the designed system(s), and eventually carrying out the desired simulation-based experimentation. All of these mentioned steps are presented orderly in this subsection.

Electrical distribution network

The existing IEEE low voltage test feeder (LVTF) distribution network is used to model the electrical distribution network [159]. There are 55 household load buses present in the IEEE LVTF model. To increase the complexity of the simulation-based experimentation in terms of size, the studied distribution network is modeled to include three districts. Each district is further divided into three sub-districts. Each sub-district is modeled as the IEEE LVTF [159]. Thus, there are a total of nine sub-districts in the system. The single-line diagram of the studied distribution network is shown in Figure 3.1.

As it can be seen in Figure 3.1, districts in the modeled distribution network are connected to the external grid through 132/33 kV grid transformers. Furthermore, each district is also connected to its sub-districts through a 32/11 kV transformer. The IEEE LVTF consists of 55 load buses [159]. An electric vehicle (EV) and PV connections are made on each load bus. There are a total of 9 sub-districts. Thus, 495 household loads, 495 EV loads, and 495 PV sources exist in the studied distribution network. The type and model of each electrical line in a sub-district are provided with the IEEE LVTF model [159]. The voltage magnitude is set to 1 per-unit at the grid bus. The external grid represents the slack bus, therefore the required instantaneous power from the external grid depends on the production/consumption in the distribution network. Thus, the power values of the external grid have not been fixed to any specific number. Also, this study has considered that the complete distribution comes under the

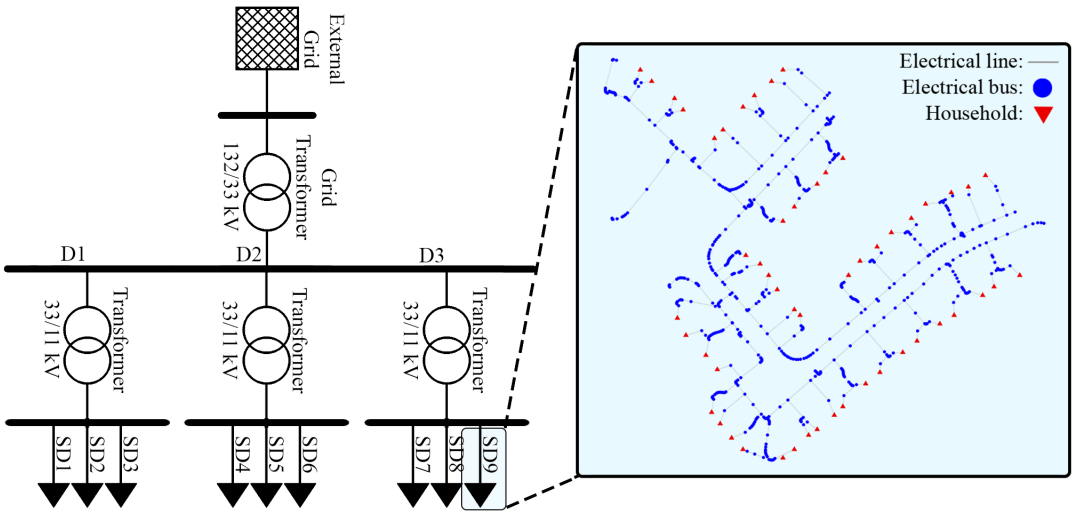

Figure 3.1: Modeled distribution network to perform simulation-based experimentation.

balance perimeter of a single BRP, although it is not a required constraint for the proposed AMAS to function (the proposed AMAS is fully capable of functioning in an environment with multiple BRPs).

Parameters

The total length of the performed simulation study is one hour. The length of each BRP imbalance settlement period is set to 15 minutes. Thus, the performance of each charging strategy during four BRP imbalance settlement periods of 15 minutes will be studied. The decision time resolution of the designed AMAS is one second i.e., each EV decides its instantaneous charging power every second. The decision time resolution is also one second for the uncontrolled charging strategy. In the case of centralized MILP optimization, the decision time resolution is set to one minute. The temporal resolution of MILP is set to one minute (and not one second) because the accuracy of the obtained solution starts saturating for a temporal resolution lower than 5 minutes while the computing time explodes. This is evident in Figure 3.2. The relationship of centralized MILP's temporal resolution with the optimality gap and the total number of decision (optimization) variables is plotted in Figure 3.2. The optimality gap is calculated by taking the obtained solution with a one-hour temporal resolution as the maximum value and the solution with a one-minute temporal resolution as the minimum value. To obtain the normalized number of decision variables curve, its values at one-hour and one-minute temporal resolutions are considered minimum and maximum values. It can be seen in Figure 3.2 that the reduction in the optimality gap is minimal when a temporal resolution of under five minutes is selected. However, the number of decision variables increases significantly. The increase in the system's complexity (number of decision variables) outweighs the improvement in the system's accuracy. Thus, MILP's temporal resolution is set to one minute, and the solution obtained with this temporal resolution is termed the optimal solution in this study.

The maximum amount of voltage deviation allowed at each bus is set to 5% of the nominal per-unit value. The rated current of each line is set based on its type, which comes along with the IEEE LVTF model [159]. The minimum and maximum allowed SoC are set to 0.3 and 0.8, respectively. This range is selected to help minimize

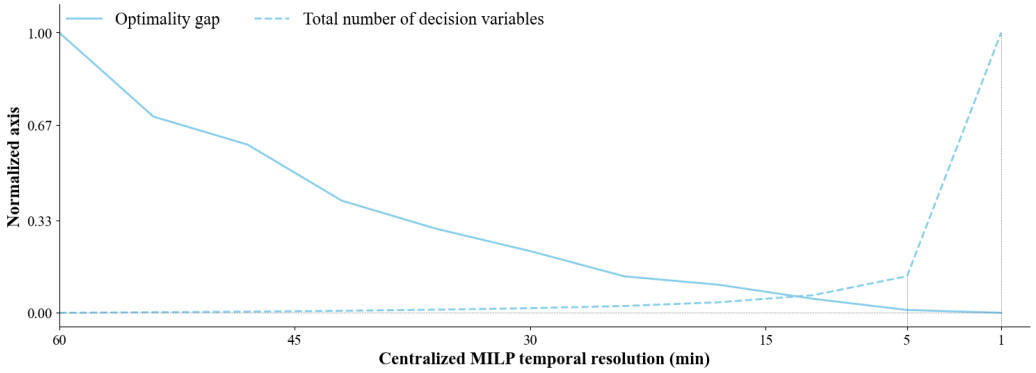

Figure 3.2: MILP temporal resolution relationship with optimality gap and total number of decision variables.

electrolyte battery degradation [55]. The desired SoC at the departure of an EV is set to 0.8 for all EVs. The rated charging power of each EV is set to 7 kW, and the battery capacity of each EV is set to 30 kWh. Each EV's charging/discharging efficiency is set to 0.95. These values are set based on the electric vehicles' specifications currently available in the market (at the time when this study was performed). In the proposed AMAS, memory-based tuning parameters k_l and k_b are both set to 0.01. The α, β, $\gamma_{l,min}$, and $\gamma_{l,max}$ to calculate the instantaneous h-value of an EV agent in the proposed AMAS are set to 0.9, 0.1, 0.1, and 0.9, respectively. A summary of these simulation parameter values is given in Table 3.1.

Datasets

Various data are required to model the functioning of different consumption/production distribution network elements during the simulation. The required *solar irradiance* data is obtained from the National Renewable Energy Laboratory (NREL) dataset [133]. The time resolution of this data is 1 minute. This data is required in the desired study to model the real-time production of PV panels present in the studied smart grid. The instantaneous PV production $P_{PV}(t)$ is calculated using the instantaneous solar irradiance $Irr(t)$ as follows:

$$P_{PV}(t) = Irr(t) A \eta_{PV} \tag{3.14}$$

In Equation (3.14), A indicates the PV panel's area and η_{PV} represents its efficiency.

Simulation parameter	Parameter value
$SoC_{e,a,min}$	0.3
$SoC_{e,a,max}$	0.8
$SoC_{e,a,depart}$	0.8
$P_{e,a,max}$	7
$P_{e,a,min}$	-7
$\eta_{e,a}$	0.95
k_l	0.01
k_b	0.01
α	0.9
β	0.1
$\gamma_{l,min}$	0.1
$\gamma_{l,max}$	0.9

Table 3.1: Values of different parameters used in the presented simulation case studies.

A forecasted PV energy production profile is also required by a BRP to calculate its day-ahead planned consumption/production $\tilde{P}(N)$, as given in Equation (2.2). Generally, forecasts for PV are provided as one-hour average values. Thus, in this study, the forecasted total PV energy production is also assumed to have a temporal resolution of one hour. The error in PV forecast is set to 10% here (approximately equal to the average of reported forecasting errors in Figure 2.2). It should be noted that it is possible that a better forecasting strategy would result in an error lower than 10%. However, designing a state-of-the-art PV forecasting strategy lies beyond the scope of this thesis. As an error-free forecast is highly unlikely irrespective of the technology utilized, this thesis focuses on designing a control strategy that would minimize the effects of forecasting errors on a smart grid in real-time. Moreover, the impact of PV energy production forecasting error on the system's performance is also studied in pseudo-stochastic studies, presented in the next section. During the simulation time, the total real-time PV production in the studied distribution network and the forecasted PV production are shown in Figure 3.3. The BRP utilizes the forecasted profile in Figure 3.3 to calculate its $\tilde{P}(N)$ during all four studied imbalance settlement periods. It is observable in Figure 3.3 that mismatches between real-time and forecasted PV profiles exist during all four studied imbalance settlement periods. During periods 1 and 2, there is over-production in the network. During periods 3 and 4, the studied distribution network is under-producing PV energy compared to the forecasted value. The BRP will face commitment mismatches in this study and require real-time optimization strategies to minimize the mismatches.

The *household* load data also needs to be modeled for each household. The IEEE LVTF model is provided with time series load profiles for the modeled households [159]. The temporal resolution of this data is also one minute. The provided load profiles are utilized in this study. To perform the desired case study, a forecasted total household load profile is also required. This forecasted time series data is used by a BRP agent to calculate its day-ahead planned consumption/production $\tilde{P}_(N)$, as given in Equation (2.2). Generally, smart meters can provide measurements at a time resolution of 10 minutes. Thus, the forecasted load profile is also assumed to have a time resolution of 10 minutes. Similar to the PV dataset, the household load forecasting error is set to 10% in this study. Again, the objective here is to minimize the impact of forecasting errors on the system in real-time. In Figure 3.4, the total real-time house-

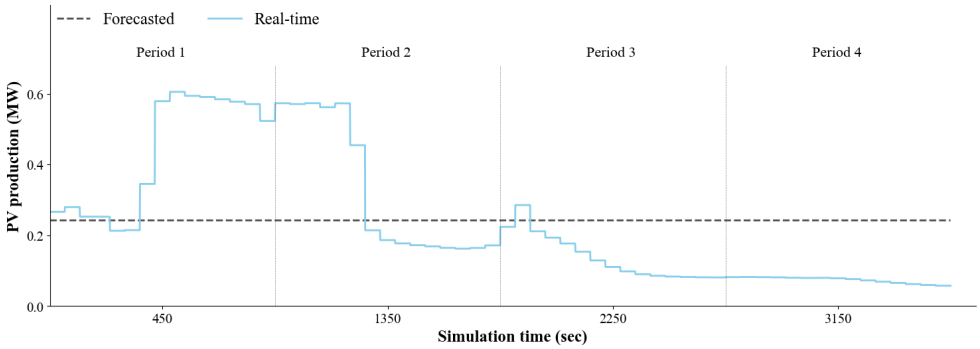

Figure 3.3: Total real-time and forecasted PV production in the studied distribution network during the simulation time.

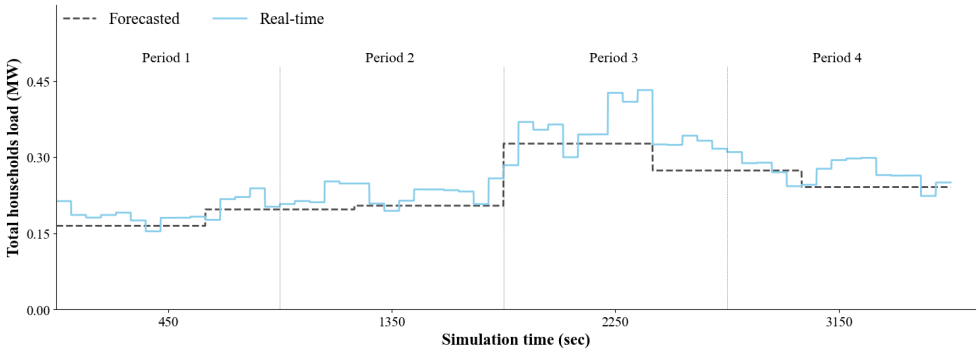

Figure 3.4: Total real-time and forecasted households load consumption in the studied distribution network during the simulation time.

hold load consumption and its forecasted value during the studied simulation time are shown. The forecasted total household load consumption profile is utilized by the BRP to calculate its day-ahead commitment value.

The presented simulation study is also designed to encounter antagonistic scenarios. This raises the complexity of the optimization problem under consideration. An antagonistic scenario is a scenario in which two different entities desire instantaneous outcomes which are opposite in nature. As shown in Figure 3.3, there is an over-production of PV energy in the balance perimeter. Thus, a BRP would desire EVs in its perimeter to increase their instantaneous charging powers (or decrease their instantaneous discharging powers). However, this could lead to network instability (due to high import current or under-voltage issues). Therefore, a critical electrical line or electrical bus would want EVs to decrease instantaneous their charging powers (or increase their instantaneous discharging powers). Thus, an antagonistic scenario will arise in the system. Sub-district *SD1* in Figure 3.1 is designed to encounter an antagonistic scenario during the studied imbalance settlement period 1. The percentage of EVs (compared to the total EV connections) in sub-district *SD1* is plotted against the simulation time in Figure 3.5. It can be seen that during period 1 of the simulation study, a larger percentage of EVs are present in the sub-district. The BRP could request all EVs for cooperation when PV energy over-production occurs. If many EVs start cooperating with the BRP by charging more (or discharging less), the network can become unstable. Thus, real-time optimization strategies would be required to handle such situations.

Implementation

The proposed adaptive multi-agent system in Section 2.4 is implemented in JAVA and is called *ADEMIS* (ADaptive Energy Management in Smart grids) [8]. The implementation is done using the AMAK framework developed by researchers at the IRIT laboratory [143]. The designed system involves co-simulation. The functionality of each agent type is defined in JAVA, and load flows are executed externally. The AMAS in JAVA utilizes a Python script to communicate with an external load flow simulator [181]. The communication is required by the AMAS platform in JAVA to set variables and launch power flow simulations in the external simulator. The described system im-

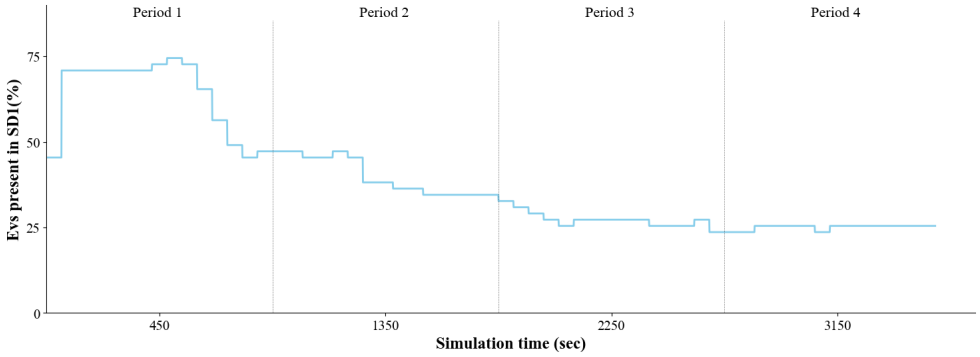

Figure 3.5: Percentage of EVs present in the sub-district *SD1* against the simulation time.

plementation is shown in Figure 3.6. At the beginning of each iteration, EV agents in the ADEMIS platform calculate their instantaneous charging/discharging power and communicate it to the external load flow simulator through the Python script. The external simulator performs the power flow and communicates the required results (electrical current, voltage, SoC, and SoH) to the agents in the ADEMIS platform. The agent cycle (iteration) ends here.

In Figure 3.7., the expanded block diagram of the designed adaptive multi-agent system is presented. The ADEMIS platform consists of the implementation of each agent type. DIgSILENT PowerFactory is the new element in the expanded system block diagram. DIgSILENT PowerFactory is a power system analysis software used in the designed system to perform power flow during each iteration [52]. DIgSILENT PowerFactory communicates currents flowing through all electrical lines, voltages at all electrical buses, state of charge of all EVs, and state of health to the ADEMIS platform. Agents in the ADEMIS platform utilize the communicated information to calculate the instantaneous charging/discharging power of each EV for the next agent cycle.

The centralized MILP optimization system is implemented in Python [181]. A Python-embedded modeling language for convex optimization, i.e., CVXPY, is used to solve and obtain the centralized MILP solutions [51]. As stated earlier, the centralized MILP solution is considered the lower bound to evaluate the performance of the designed adaptive multi-agent system. It can be argued that the stated MILP formulation involves approximations and linearizations. These approximations can impact the accuracy of the obtained solution when applied in real-life. Therefore, to evaluate the impact of the approximations on the solution's quality, the optimized EV power profiles obtained through the centralized MILP strategy are imported and simulated in DIgSILENT PowerFactory using the studied distribution network. Thus, a comparison is made between the results obtained through MILP optimization and the results

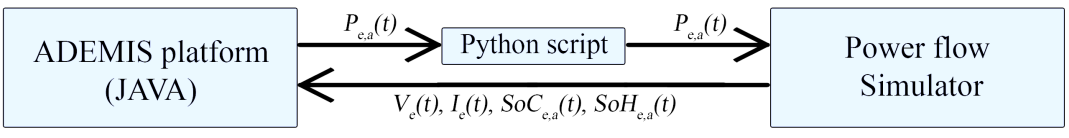

Figure 3.6: Block diagram of the implemented AMAS for smart grid energy optimization.

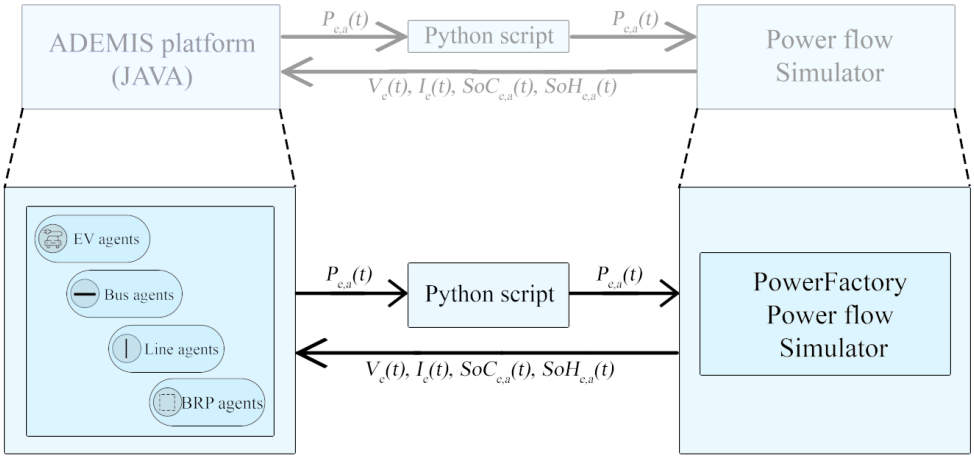

Figure 3.7: Expanded block diagram of the implemented AMAS for smart grid energy optimization.

obtained through simulating MILP optimization results in DIgSILENT PowerFactory. The calculated percentage cosine similarity between the active power grid profiles of MILP optimization results and MILP optimization results simulated in DIgSILENT PowerFactory comes out to be 99.02%. It confirms a minor impact of MILP approximations on the quality of the obtained solution and, thus, making MILP optimization results comparable to the other two studied strategies (i.e., uncontrolled and AMAS).

Computing machine's specifications

Specifications of the computing machine used to perform the designed simulation case study are listed in Table 3.2:

Evaluation metrics

It is important to describe the metrics considered to evaluate the performance of the proposed adaptive multi-agent system. In this simulation case study, the following metrics are included:

- Constraints satisfaction
- Optimality

	Specification
Processor	Intel Core i5-10210U (1.6 GHz)
Memory	DDR4 (8 GB, 2666 MHz)
Storage	SSD (256 GB, 7300 MB/s)

Table 3.2: Specifications of the computing machine used to perform simulation-based experiments.

- Scalability

Each of the stated metrics is described below in detail.

Constraints satisfaction

Constraints stated in Section 2.1 are expected to be satisfied by the optimization-based strategies. If an uncontrolled strategy is followed, these constraints are more likely to be violated. On the other hand, these constraints should be satisfied as the hard constraints of the centralized MILP formulation. Thus, the designed adaptive multi-agent system will be compared with uncontrolled and centralized MILP optimization strategies. The evaluation will be based on whether all desired constraints are satisfied.

Optimality

The objective of the studied optimization problem is to minimize BRP's commitment mismatch, as stated in Equation (2.1). The proposed AMAS will be evaluated against the stated two baseline strategies regarding its optimality. The uncontrolled baseline strategy is considered the upper bound, while the centralized MILP optimization strategy is considered the lower bound. Ideally, the solution obtained through the proposed adaptive multi-agent system strategy should be near-optimal (close to the lower bound), if not optimal.

Scalability

The implemented optimization strategy should be able to optimize a larger-scale smart grid in real-time. An optimization strategy can be optimal, but its real-life implementation becomes a question if it is not scalable. Therefore, it is crucial to draw a comparison between the centralized MILP baseline strategy and the proposed adaptive multi-agent system optimization strategy. The scalability comparison between both strategies is made in terms of their practical computational requirements (i.e., required computation time and memory) [77].

Note 3.2.1

The *computational requirements* of an algorithm can be defined as the amount of computing resources required to execute it. Both time and memory are generally considered the most prominent resources required to execute an algorithm. Time resource is the time an algorithm takes to complete its execution, and memory resource is defined as the total memory required to execute an algorithm.

In this study, a smart grid optimization algorithm is termed as scalable if it manages to optimize a large-scale network in real-time (i.e., it does not require a large amount of time for its execution) and is not memory-intensive (i.e., it does not require a large amount of memory for its execution). Ideally, both time and memory computational

requirements should be independent of the size of the studied smart grid, and they should not increase with the size of the distribution network.

Results

In this subsection, a comprehensive analysis of the simulation case study results is provided, highlighting the performance of the proposed adaptive multi-agent system and the previously discussed baseline strategies.

Constraints satisfaction

As stated earlier, the distribution network in the designed case study faces a variety of challenges. These challenges also include network instability due to congestion in an electrical line or voltage limit violations at an electrical bus. In Figure 3.8, a comparison of *current* flowing through the line connecting sub-district *SD1* with district *D1* is presented. The presented electrical current curves are obtained using the proposed adaptive multi-agent system strategy and two baseline strategies, i.e., uncontrolled and centralized MILP optimization strategies.

An electrical line congestion can be observed in Figure 3.8, during the imbalance settlement period 1. Electrical line congestion is observed if the uncontrolled charging strategy is followed. This electrical line congestion occurs due to a large percentage of EVs present in the sub-district *SD1* during the imbalance settlement period 1, as shown in Figure 3.5. These EVs charge at their maximum power and cause line congestion in the network when they remain uncontrolled. The observed electrical line congestion lasts for 30.16% of the total imbalance settlement period. The average value of the observed line congestion is 110.50% of the rated current value. This significant amount of electrical line congestion can make the distribution network unstable. Therefore, a control strategy is desired to prevent electrical line congestion from arising in the uncontrolled strategy.

In the centralized MILP optimization strategy, the prevention of electrical line congestion is a hard constraint. Thus, the obtained solution does not result in an electrical line congestion, as shown in Figure 3.8. In the proposed adaptive multi-agent system, there is no concept of hard constraints. Rather, agents cooperate with each other to

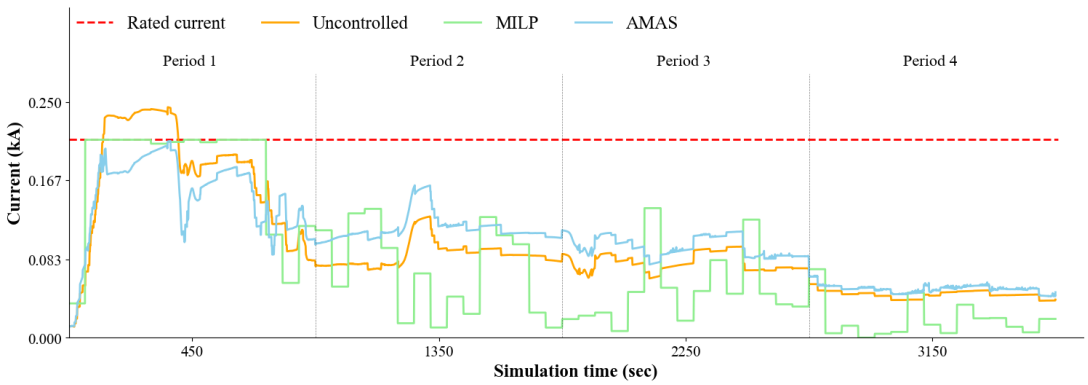

Figure 3.8: Comparison of electrical currents during the simulation time.

achieve the global objective(s) of the system. As shown earlier in Figure 3.3, there is an over-production of PV energy compared to its forecasted value during imbalance settlement period 1. Thus, the BRP would want EVs to charge more (or discharge less). On the other hand, the line agent would desire EVs to charge less (or to discharge more) to prevent line congestion. Thus, an antagonistic situation arises, which makes the decision-making process of an EV more complex. Ideally, EVs should cooperate with both antagonist agents and thus prevent line congestion from occurring. In Figure 3.8, it can be observed that no electrical line congestion occurs when the proposed AMAS strategy is followed. It means that the EV agents cooperate with critical line agents in the system to prevent electrical line congestion from happening.

Another network problem that can make a distribution network unstable (thus violating the DSO constraints) is the *voltage* deviation from its nominal value beyond the allowed limit. The voltage magnitude results obtained at the last bus of the sub-district *SD1* during the simulation study are presented in Figure 3.9. During the imbalance settlement period 1, voltage constraint violation can be observed if the uncontrolled charging strategy is followed. The reason is a large number of EVs charging simultaneously, which causes an under-voltage issue in the distribution network. The observed voltage constraint violation lasts for 28.37% of the total imbalance settlement period. The voltage magnitude is, on average, 0.21% lower than its allowed voltage limit when the voltage constraint is violated. In the centralized MILP optimization, no voltage constraint violation is observed as it is a hard constraint of the optimization problem, Equation (2.11). In the proposed adaptive multi-agent system platform, EV agents have to cooperate with both the BRP agent and the bus agent during period 1. The BRP agent requires EVs to charge more (or to discharge less) to minimize its energy mismatch occurring due to PV energy over-production during period 1, as shown in Figure (3.3). Whereas, the bus agent wants EVs to charge less (or discharge more) to prevent under-voltage issues happening at its electrical bus. In Figure 3.9, no voltage constraint violation is observed for the proposed AMAS. Therefore, it can be confirmed that the designed AMAS functions as expected. It manages to minimize the criticality of a critical bus agent through cooperation, even under an antagonistic situation. Prosumer constraints, Equation (2.14)−(2.17), are satisfied in the case of all studied strategies. All electric vehicles manage to acquire the desired SoC at their respective departure times.

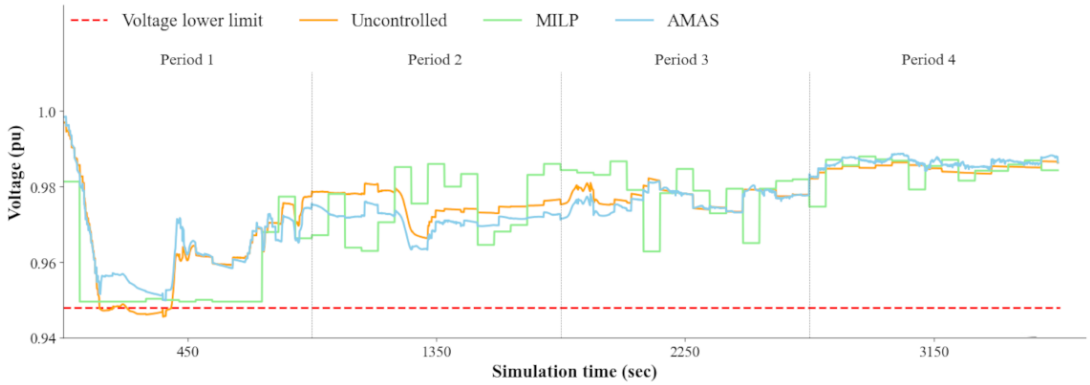

Figure 3.9: Comparison of electrical voltages during the simulation time.

Optimality

Minimizing the BRP commitment mismatch is the objective of the studied smart grid optimization problem. There are four imbalance settlement periods in the studied simulation case study. The obtained BRP commitment mismatch results for each studied strategy are shown in Table 3.3. The centralized MILP optimization strategy manages to completely minimize the BRP commitment mismatch. Thus, this solution is used as the lower bound to evaluate the performance of the uncontrolled strategy and the designed adaptive multi-agent system optimization strategy.

In the uncontrolled strategy, EVs are allowed to charge at their rated power, without any constraints, at all times. The BRP cannot control EVs to minimize its energy mismatch in the uncontrolled strategy. Thus, a significant commitment mismatch is observed when the uncontrolled strategy is followed in Table 3.3. On average, a commitment mismatch of 33.58% is observed per imbalance settlement period in the performed simulation study. This value can be considered as an upper bound. Any proposed smart grid optimization strategy is desired to be below this upper bound and be close to the MILP lower bound as much as possible. In the proposed AMAS optimization strategy, the BRP agent has to request cooperation from all EV agents. It is seen that EV agents managed to cooperate with line and bus agents and maintained the stability of the distribution network, Figure 3.8 and Figure 3.9. Through the results stated in Table 3.3, it is also confirmed that EV agents also managed to cooperate with the BRP agent as a significant reduction in the commitment mismatch values can be observed compared to the uncontrolled strategy. The proposed AMAS optimization strategy managed to reduce the commitment mismatch by 99.5% compared to the uncontrolled strategy (upper bound). The AMAS solution is near-optimal as it is close to the optimal MILP optimization solution (lower bound).

Both optimization strategies (MILP and AMAS) manage to minimize the BRP commitment mismatch by controlling the instantaneous charging/discharging power of each EV in real-time. The sum of instantaneous charging/discharging powers of all EVs in the studied distribution network during the simulation time is shown in Figure 3.10. In Figure 3.10, during the initial two imbalance settlement periods, EVs are consuming more power compared to the forecasted (planned) instantaneous power consumption. This is due to the fact that PV panels are producing more energy than ex-

	Strategy		
	Uncontrolled	**AMAS**	**MILP**
Period 1 (kWh)	38.06	0.02	0
Period 2 (kWh)	11.13	0.01	0
Period 3 (kWh)	37.20	0	0
Period 4 (kWh)	47.93	0.63	0
Total (kWh)	134.32	0.66	0

Table 3.3: Commitment mismatch comparison between studied strategies.

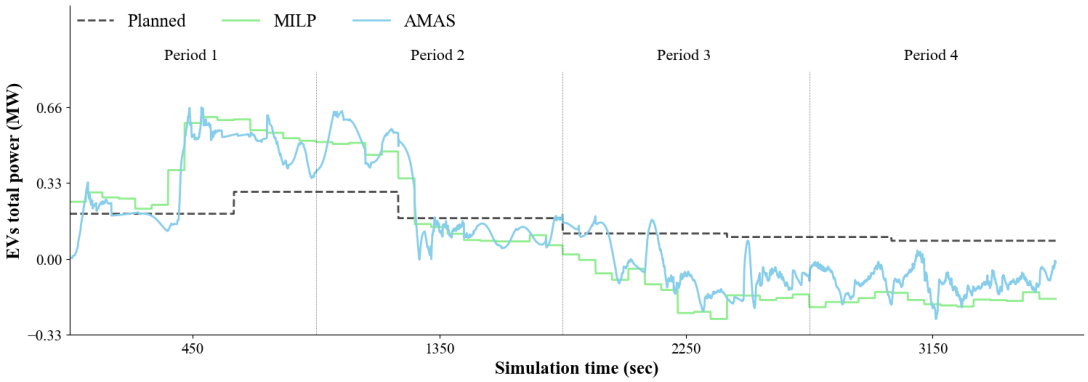

Figure 3.10: Comparison of electrical vehicles' total instantaneous powers during the simulation time.

pected during these two periods, as presented in Figure 3.3. In the third and the fourth BRP imbalance settlement periods, EVs' total consumption is lower than planned. This is due to the under-production of PV energy during these imbalance settlement periods. Unlike the centralized MILP optimization (which is fully deterministic here), the proposed AMAS does not require aggregated input data in order to optimize the system, nor it is assumed to know the future. Instead, the proposed AMAS is purely reactive in nature. Electric vehicle agents do not make any predictions regarding the future states and only react to the current state of the distribution network. However, as it can be observed in Figure 3.3, the solution profile of AMAS still manages to follow the path obtained by the deterministic MILP optimization solution for the total consumption of all EVs. It is a significant result as it confirms that the proposed AMAS can produce near-optimal solutions only through its reactive approach without knowing any future information. It should be noted that the oscillations through rapid charging and discharging of an electric vehicle's battery have not been penalized in the proposed AMAS system and thus it may lead to faster battery degradation. To tackle this problem a more sophisticated EV battery charging/discharging model along with constraints on rapid charging and discharging of an EV battery can be designed. These improvements will be a part of the future works.

Scalability

Comparisons of constraints' satisfaction and optimality have put the proposed AMAS strategy over the uncontrolled strategy in terms of optimizing the studied smart grid. However, the MILP optimization strategy has performed better than the proposed AMAS strategy in terms of optimality. But, in a real-life scenario, other dimensions than just optimality must be considered when selecting an optimization strategy. This dimension is the scalability of the selected optimization strategy. As in a practical smart grid, one can expect the total number of EVs to be in the thousands, if not millions. Thus, the deployed optimization strategy should demonstrate scalability by being able to optimize a large-scale smart grid.

A comparison of practical computational requirements is made in Figure 3.11. The comparison is made both in terms of computation time and memory requirements. The

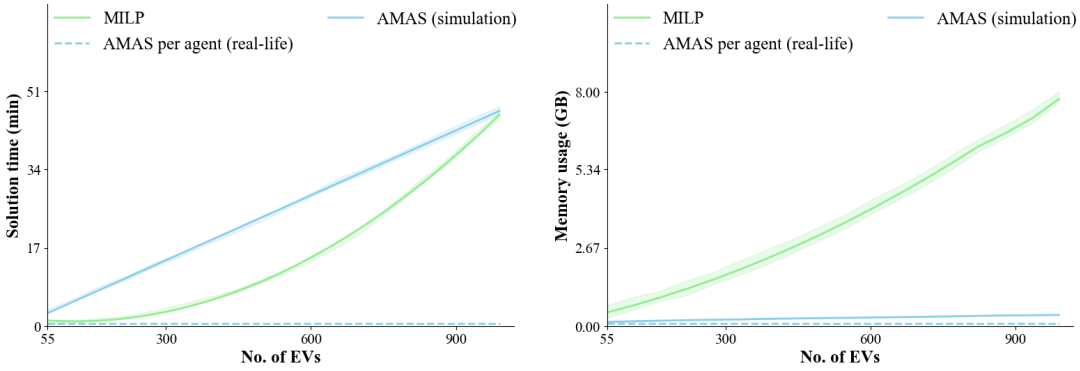

Figure 3.11: Comparison of computation time (left) and memory requirements (right).

presented curves are obtained by simulating each optimization strategy ten times for each x-axis data point (i.e., the number of EVs in the network). In Figure 3.11, two types of practical computational requirements assessments are reported for the AMAS. Although the simulation study is performed on a single computing machine, the AMAS system is designed to be fully decentralized in a real-life (i.e., deployment) scenario. It means each agent is supposed to have its own computing power, which would lower the practical computational requirements of the AMAS through decentralization.

In terms of computation time requirements, the centralized MILP strategy performs better than the proposed AMAS strategy to carry out simulation studies when the total number of EV agents is lower than 1000. However, the centralized MILP is combinatorial in nature and belongs to the *NP* (nondeterministic polynomial time) class of optimization problems. The combinatorial nature of the centralized MILP results in its inability to scale well. As seen in Figure 3.11, the centralized MILP optimization time increases quadratically with the number of total EVs in the system. Thus, the studied centralized MILP optimization strategy may perform well on a small-scale system with a few hundred EVs. But, it will not be able to produce real-time solutions when applied to a large-scale smart grid. On the other hand, the optimization time increases linearly when the proposed AMAS is utilized to perform simulation studies. Thus, after 1000 EVs, the AMAS is outperforming the studied MILP strategy to carry out simulations. The advantage of the proposed AMAS is even more apparent when a real-life situation is assumed. In a real-life scenario, the centralized MILP would still perform optimization on a single node. Whereas, a practical AMAS will utilize decentralized computing. Furthermore, the designed AMAS is reactive, and each agent only requires milliseconds to complete its one-agent cycle. That is why the proposed AMAS is always outperforming the centralized MILP, in Figure 3.11. The computation time of each agent type in AMAS is shown in Table 3.4. This is the expected time taken by each agent type to complete its one agent cycle. The maximum time taken by any agent type in Table 3.4 is in milliseconds. Thus, the designed AMAS is expected to operate in real-time on a large-scale smart grid as the reported computation times are independent of the size of the total number of agents in the system.

Memory complexities of the studied optimization strategies are also presented in Figure 3.11. It can be seen that the AMAS is clearly outperforming the MILP strategy in terms of the required memory. This is because the MILP strategy needs to process a

Agent type	Computation time per agent cycle (ms)
Line	0.003
Bus	0.004
BRP	0.064
EV	0.540

Table 3.4: Computation time of each AMAS agent type to complete one agent cycle.

large amount of input data to perform optimization. Whereas, the designed AMAS is reactive and thus only requires input data corresponding to the current instant. That is why a significant difference between the memory requirements of both optimization methods can be seen in Figure 3.11. Based on the comparison of practical computational requirements, it can be concluded that the centralized MILP strategy can be a good option for conducting small-scale simulation case studies. But if the intent is to find a scalable optimization strategy (i.e., suitable for real-life implementation or to simulate large-scale EV fleets), then the designed AMAS system is clearly outperforming the MILP optimization strategy. It must be noted that a hierarchical system based on MILP can be designed. However, such systems may still suffer from scalability challenges as discussed earlier in Section 2.2. Such systems have not been modeled and studied here as that was not the focus of this thesis.

Note 3.2.2

It is important to mention that it is not a functional requirement for each line/bus to have its sensor for the developed methodologies to work (i.e., estimations could be performed as well). In addition, power/voltage measurements collected by currently installed sensors (e.g. smart meters) could be used to limit the need for additional sensors. Also, the system operator can install sensors at key interest points in the network where congestion is more probable and thus it will limit the number of total sensors required in the system (i.e., total cost of the system).

3.3 Pseudo-stochastic simulation-based experimentation

Introduction to pseudo-stochasticity

In the previous section, it was observed that the proposed AMAS produces a near-optimal solution. However, in the performed simulation study, no stochasticity was considered. Whereas in real-life scenarios, conditions are stochastic (e.g., stochasticity in forecasted PV irradiance, or in load consumption etc.) [73]. Hence, it is crucial to

study the impact of stochasticity on the proposed AMAS, which would help determine if the designed AMAS can perform under real-life conditions. This would improve the performance of AMAS under uncertain conditions. In this section, the impact of uncertainties (stochasticity) in the forecasted PV energy production is studied. Only one type of stochasticity (i.e., PV energy production) is studied here because the objective is to find if there is a significant degradation in the performance of the designed AMAS under stochasticity, and not to see the impact of each type of stochasticity on the designed AMAS. If a significant degradation in the designed AMAS is observed, then it should be improved (using novel approaches such as machine learning that would lead to self-organization). Furthermore, instead of performing a stochastic analysis immediately, a relatively simpler pseudo-stochastic analysis is performed on the designed AMAS. If a reduction in the system's quality is observed in the pseudo-stochastic study, then one can directly work towards improving the system's performance under uncertainty. Otherwise, one can move towards performing stochastic studies if no impact of pseudo-stochasticity is observed on the designed system.

Here, the impact of uncertainty in the PV irradiance forecast error is studied. This error is directly linked to the forecasted PV energy production profile. *Stochasticity* (or *uncertainty*) in a variable is associated with randomness. A *random variable* is generally described as a variable whose possible outcomes are random (or non-deterministic) in nature. The value of this variable is non-deterministic, and each possible outcome of this variable is associated with a probability. As stated above, instead of stochasticity, the impact of pseudo-stochasticity on the designed AMAS is studied in this section. In contrast to stochasticity, a variable's *pseudo-stochasticity* involves pseudo-randomness. A *pseudo-random* variable is a variable that appears to be random but, in reality, is not. In fact, it is generated through a deterministic and repeatable process.

Persistence model for pseudo-stochastic scenarios generation

The PV forecasting error was set to 10% in the simulation study of the previous section. In this section, the PV irradiance forecasting error has not been set to any specific value. Instead, it is modeled as a pseudo-random variable here. Thus, this variable can have different values, and the impact on the designed AMAS can be observed for each value. The persistence algorithm approach is used to generate values of this pseudo-random variable (i.e., PV irradiance forecasting error). A *persistence algorithm* is a naive time series forecasting algorithm that assumes the system remains unchanged. Thus the values of the time series persist between the present and the future, i.e., predicted time series values are equal to present time series values. The persistence model to predict solar irradiance data can be defined as follows:

Persistence solar irradiance prediction model

$$Irr_p(t+h) = Irr_r(t) \tag{3.15}$$

In Equation (3.15), h represents the prediction horizon. Thus, this equation states

that the predicted irradiance value at some future instant $Irr_p(t+h)$ is equal to the real irradiance value at the present instant $Irr_r(t)$. However, the goal is not to generate only one solar irradiance prediction profile through the persistence model, but to generate a number of pseudo-random persistence model profiles depending on the error that the persistence prediction model may encounter. Thus, the following *pseudo-stochastic persistence solar irradiance prediction model* is formulated:

Pseudo-stochastic persistence solar irradiance prediction model

$$\begin{aligned} Irr_p(t+h)_n &= Irr_r(t) + Irr_r(t)e_f(t) \\ &= Irr_r(t) + Irr_r(t)\left(\varrho e_f(t-1) + \eta\right) \end{aligned} \tag{3.16}$$

In Equation (3.16), $Irr_p(t+h)_n$ stands for the predicted irradiance value at time $(t+h)$ for generated pseudo-stochastic irradiance profile n. Term $e_f(t)$ represents the error in the forecast. Based on different values of $e_f(t)$, one can obtain different solar irradiance prediction profiles. The value of this forecasting error $e_f(t)$ is linked to its value at the previous instant $e_f(t-1)$. This dependency is due to the periodicity present in this time series forecasting error. The relationship between $e_f(t)$ and $e_f(t-1)$ can be found through auto-correlation. *Auto-correlation* is used to calculate the correlation (similarity) between a time series signal and a delayed copy of itself. Thus, auto-correlation can help in identifying repeating patterns in a time series. In Equation (3.16), ϱ represents the auto-correlation coefficient of the forecasting error with a copy of itself delayed by a unit step. One can calculate $e_f(t)$ by using ϱ and $e_f(t-1)$. The values for ϱ and $e_f(t-1)$ depends on the forecaster. Thus, one can utilize past data to train and evaluate the performance of the modeled forecaster, which would result in finding ϱ and $e_f(t-1)$ for the modeled forecaster. Gaussian noise η is added to the prediction model to emulate errors that may arise due to natural sources.

Note 3.3.1

The persistence model is a naive forecaster, usually considered a baseline while developing novel forecasters. Thus, it can be argued that a better forecaster model can be used to generate pseudo-stochastic solar irradiance scenarios. However, it must be noted here that the purpose of this pseudo-stochastic study is not to model a state-of-the-art solar irradiance forecaster but to evaluate the impact of inherent forecasting errors on the designed adaptive multi-agent system's performance. Therefore, a naive persistence prediction strategy will suffice to perform the desired pseudo-stochastic study.

Simulation-based experimentation settings

Simulation-based experimentation settings are kept the same as in Section 3.2 to perform pseudo-stochastic studies, i.e., studied optimization problem, distribution network, parameters, implementations, and loads and EVs datasets are the same. The

only change is in modeling the PV energy generation profile during the simulation studies. As the forecasting error is a pseudo-random variable, different values of this variable are used to generate different PV energy production scenarios. Simulation studies are performed considering each scenario as the real-time PV energy production input. The conducted simulation studies will help analyze the impact of forecasting errors in PV energy production on the designed adaptive multi-agent system. The generated pseudo-stochastic PV energy production profiles are presented as follows.

Pseudo-stochastic scenarios generation

Solar irradiance data from the National Renewable Energy Laboratory (NREL) are used to generate pseudo-stochastic scenarios [133]. The temporal resolution of the obtained data is one minute. The time horizon h in the persistence forecaster model is set to 1440 (one day), i.e., the solar irradiance value at the present instant is the same as it was yesterday at the same exact instant. Based on this setting, the persistence solar irradiance prediction model, in Equation (3.15), is trained and evaluated on six months of NREL solar irradiance data [133]. The obtained forecasting error and the auto-correlation of this error during the evaluation phase of the forecaster are shown in Figure 3.12.

Based on the results in Figure 3.12, the pseudo-stochastic persistence solar irradiance prediction model, given in Equation (3.16), is modeled. The forecasting error's auto-correlation ϱ is set to 0.7. This value is obtained from the auto-correlation plot in Figure 3.12 as the auto-correlation of the forecasting error with unit lag (i.e., delayed by a unit step) is equal to 0.7. To determine $e_f(t)$ in Equation (3.16), the value of $e_f(t-1)$ is also required. The value of $e_f(t-1)$ is sampled from the probability density function of the observed forecasting error during the evaluation phase. Based on the sampled $e_f(t-1)$, a number of pseudo-stochastic solar irradiance scenarios are generated. These pseudo-stochastic scenarios are presented in Figure 3.13.

Figure 3.13 shows the present-day solar irradiance profile (solid line) during the simulation time. The BRP uses this irradiance profile to calculate its day-ahead production/consumption schedule. However, the real-time solar irradiance can differ from its expected value the next day due to forecasting errors. A number of pseudo-

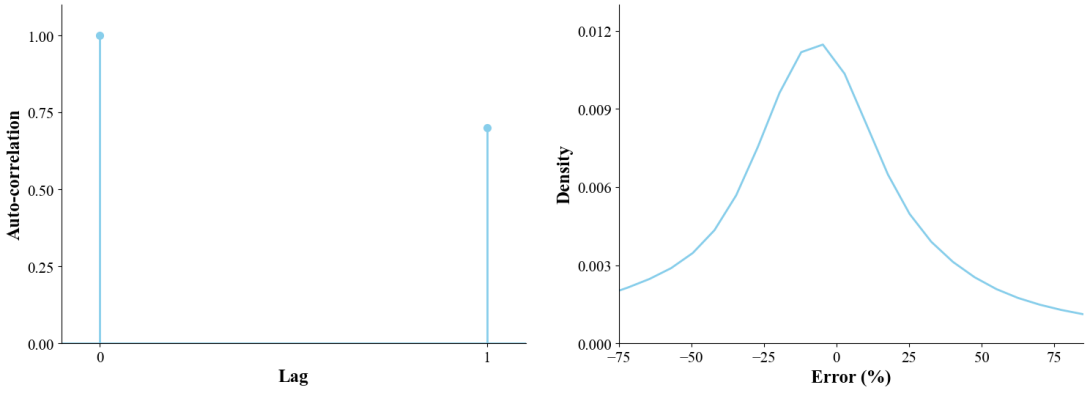

Figure 3.12: Auto-correlation of the forecaster error (left) and probability density function of the observed forecasting error (right).

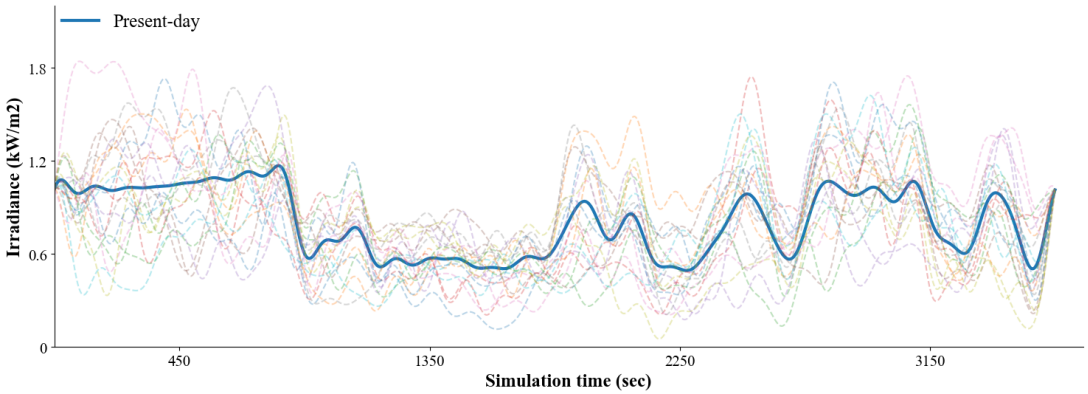

Figure 3.13: Current day solar irradiance (solid line) and forecasted pseudo-stochastic solar irradiance profiles (dotted lines).

stochastic solar irradiance scenarios are generated based on the varying amount of forecasting errors (sampled from probability density function Figure 3.12). In Figure 3.13, these pseudo-stochastic scenarios (dotted lines) show how real-time solar irradiance can vary compared to the expected solar irradiance profile (solid line) due to forecasting errors. A number of simulations are performed to analyze the impact of forecasting errors on the designed AMAS. In each simulation, one of the pseudo-stochastic scenarios (dotted lines) in Figure 3.13 is selected to model the real-time solar irradiance in the simulation. The remaining simulation conditions remain unchanged.

Evaluation metrics

The performance of the proposed adaptive multi-agent system is studied in terms of *optimality* and *constraints satisfaction* here as well. The aim is to quantify the impact of pseudo-stochasticity in PV forecasting error on the designed AMAS. The centralized MILP optimization is used to determine the optimality lower bound. It must be kept in mind that although the centralized MILP approach helps find the lower bound for each pseudo-stochastic study here, in real-life it is not stochasticity-free. This is because an error-free solar irradiance forecast is unlikely. Another essential objective is to see if the pseudo-stochasticity results in constraint violations.

Results

A total of 20 pseudo-stochastic simulations are performed. The total simulation time is one hour, i.e., four imbalance settlement periods. The obtained results of all pseudo-stochastic simulations are presented and discussed below.

Optimality

The objective during each pseudo-stochastic simulation is the same, i.e., to minimize the BRP commitment mismatch error during all four studied imbalance settlement periods. The obtained commitment mismatch values at the end of each imbalance settlement period are shown in Figure 3.14 for each pseudo-stochastic simulation study.

	% of AMAS pseudo-stochastic studies with constraints violation	% of MILP pseudo-stochastic studies with constraints violation
Electrical line congestion	50	0
Voltage limit violation	35	0
Prosumer desired SoC	0	0

Table 3.5: AMAS and MILP constraints satisfaction results of pseudo-stochastic simulation studies.

It can be seen that the lower bound (MILP solution) is equal to zero for all pseudo-stochastic studies. Thus, an optimal solution of null BRP commitment mismatch exists for each study. However, the commitment mismatch values obtained through the designed AMAS are not equal to zero (lower bound). Additionally, the commitment mismatch value obtained by the AMAS for each pseudo-stochastic study differs. The average total commitment mismatch of all 20 pseudo-stochastic studies is 133.63% higher than that of the deterministic simulation study. This analysis confirms that the optimality of the designed system can indeed be further improved under uncertainties.

Constraints satisfaction

While minimizing the BRP commitment mismatch, a set of constraints must be satisfied. This set of constraints includes DSO constraints (no electrical line congestion and voltage limits violation), and prosumers' constraints. These constraints are given in Equations (2.10), (2.11), and (2.17). A comparison of the constraints satisfaction through the designed AMAS and the MILP is presented in Table 3.5. In the case of pseudo-stochastic studies conducted using the centralized MILP optimization, there are no constraint violations. However, when the same pseudo-stochastic studies are performed using the designed AMAS platform, a significant number of constraint violations are observed (electrical line congestion in the line connecting sub-district *SD1* to its district *D1* and electrical voltage limit violation at the last bus of the sub-district *D1* in Figure 3.1). These constraints belong to the DSO's set of constraints. Thus, the stability of the network is observed to be compromised here if the designed AMAS system is utilized. Half of the pseudo-stochastic simulation studies violated the DSO line constraint. The observed electrical currents during these constraint violations are 4.76% higher than the rated current of the line on average. The voltage constraint violation is also observed in over one-third of the pseudo-stochastic simulation studies. The average voltage during the occurrence of the voltage constraint violation is 0.21% higher than its allowed limit value.

Based on the obtained optimality and the constraints satisfaction results in this sec-

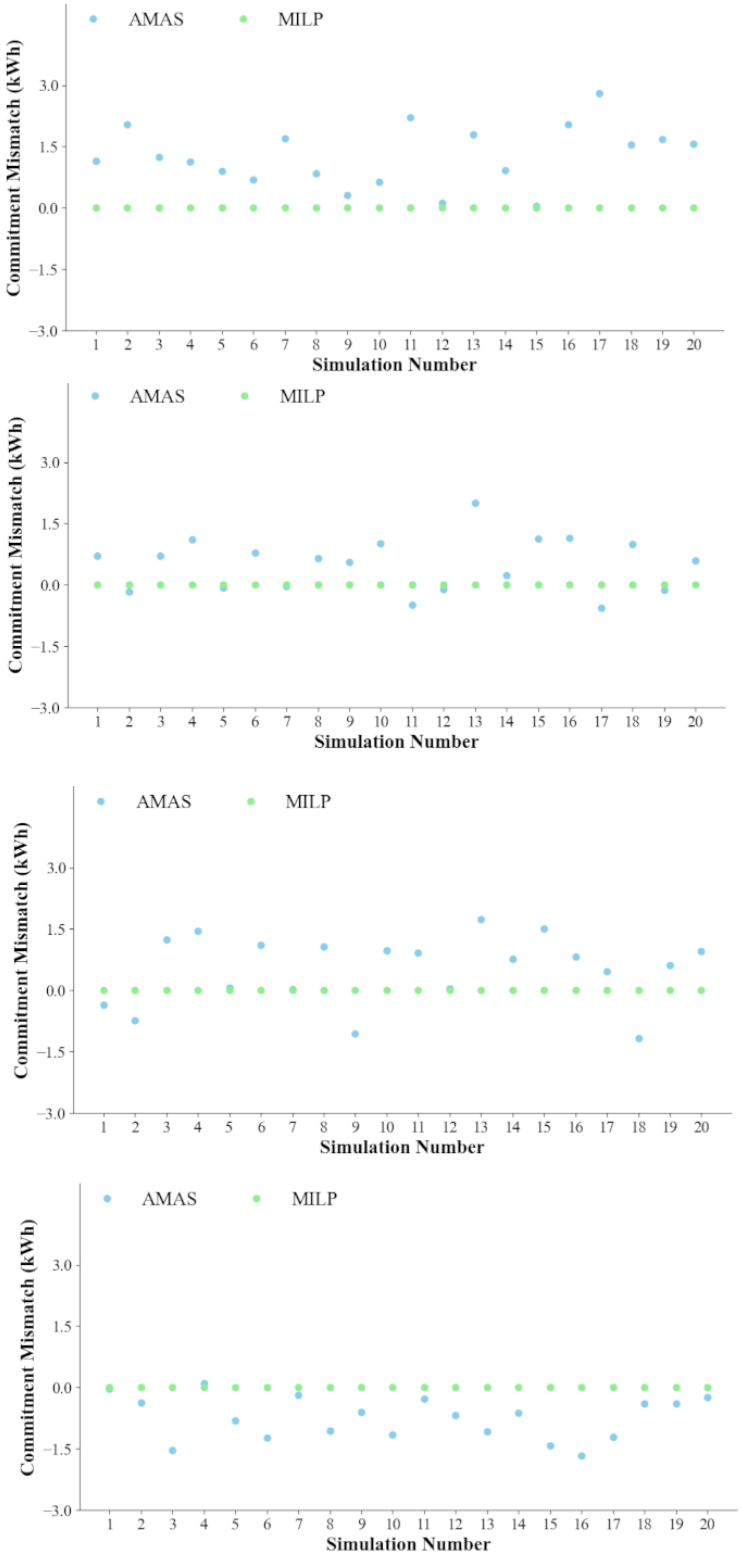

Figure 3.14: Obtained optimality (commitment mismatch) results for each pseudo-stochastic study during period 1 (top), period 2 (second from the top), period 3 (third from the top), and period 4 (bottom).

tion, it can be confirmed that pseudo-stochasticity (hence stochasticity as well) has a significant impact on the system of the designed AMAS in which agents lack anticipative abilities. This can be attributed to the designed AMAS utilizing only reactivity. Each agent in the current AMAS is only reacting to the changes in its environment. However, one can also model an AMAS agent to learn from the history of its interactions. This change will result in an intelligent (and not only reactive) agent in nature and is expected to perform better under real-life stochastic conditions than the currently proposed adaptive multi-agent system.

3.4 Conclusion

This chapter presented a detailed evaluation of the proposed adaptive multi-agent system in Chapter 2. The designed system was utilized to optimize the studied real-time grid balancing optimization problem, and the obtained results were compared with two other baseline strategies. These baseline strategies were the uncontrolled strategy and the centralized MILP optimization strategy. The obtained results are summarized in Figure 3.15. The presented Kiviat diagrams compare the performance of all three studied strategies regarding their optimality, constraint satisfaction, and practical computational requirements (solution time and memory). It can be seen in Figure 3.15 that an uncontrolled strategy results in no computational requirements (being a non-optimization strategy), but the system is not optimal. Furthermore, network instability is also observed when the uncontrolled strategy is followed, as DSO's constraints are not always satisfied. These issues are solved if the centralized MILP optimization strategy is followed. However, the MILP optimization strategy does not scale well due to high time and memory complexities. The proposed AMAS can help in achieving the best of both worlds. It manages to satisfy all the required constraints while being near-optimal. Additionally, being a decentralized system results in minimal time and memory complexities. Thus, it puts itself forward as a strong candidate to optimize practical large-scale smart grids in real-time.

The initial results obtained for the proposed AMAS were optimistic. However, the simulation study was performed under deterministic conditions, which is not the case in real-life. Thus, a pseudo-stochastic study was also designed and performed to evaluate the performance of the presented AMAS under pseudo-stochasticity. The obtained results showed that the designed AMAS does not perform as desirably due to the lack of anticipative capabilities in its agents. This negative impact could hinder the adoption of the adaptive multi-agent system from optimizing smart grids in real-life. Thus, the following two chapters focus on one of the techniques (i.e., reinforcement learning) that can be used to minimize the impact of stochasticity on the designed adaptive multi-agent system. The next chapter discusses a novel reinforcement learning-based methodology to optimize smart grids in a decentralized manner.

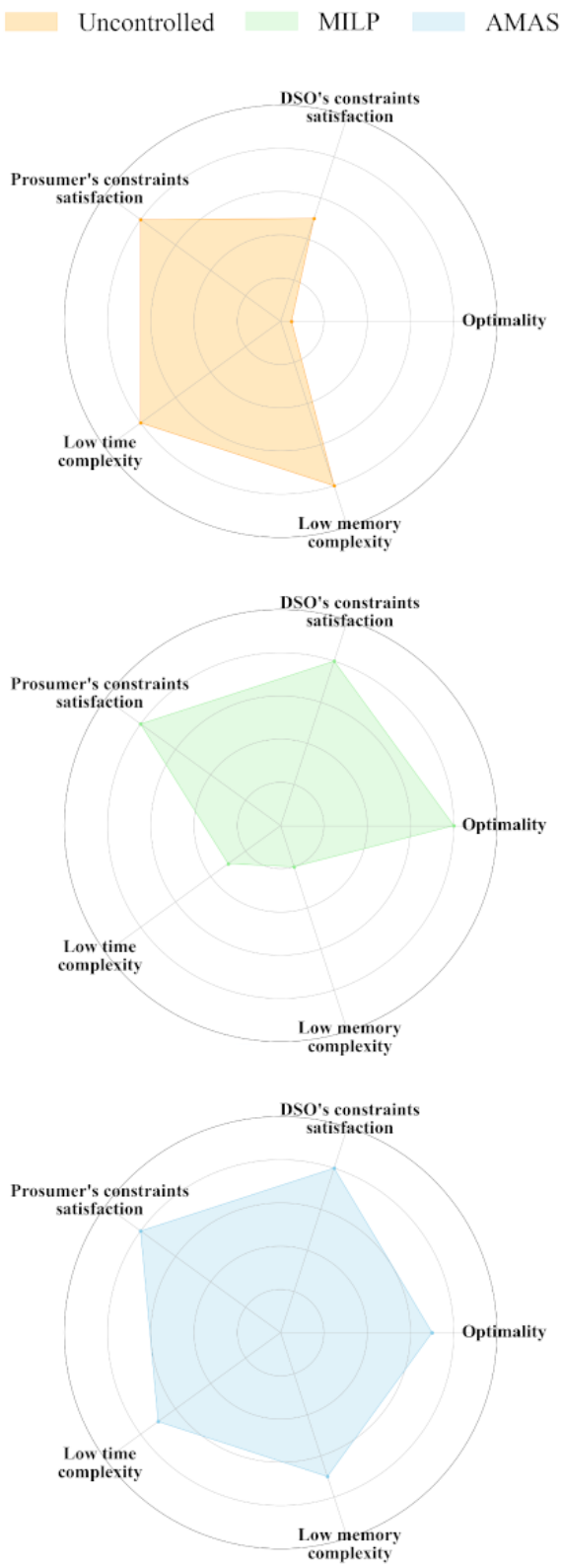

Figure 3.15: Kiviat diagrams comparing performances of the studied strategies: uncontrolled (left), MILP (middle), and AMAS (right).

Chapter 4

Decentralized multi-armed bandit for smart charging under uncertainties

> Such (easy to understand) are all truths, once they are found; but the difficulty lyeth in finding them.
>
> Galileo Galilei

Summary

In the preceding chapter, it was observed that a heuristic adaptive multi-agent system may encounter performance degradation due to the absence of anticipative abilities in the decision-making agents. It was deduced that incorporating reinforcement learning could enhance the system's performance, especially in dealing with uncertainties. This chapter progresses in that direction by introducing a fully decentralized real-time smart grid control system, where agents employ reinforcement learning to enhance solution quality under uncertainties. Specifically, the agents in this developed system will utilize combinatorial multi-armed bandit learning.

Contents

This chapter introduces the design of a reinforcement learning-based multi-agent system for real-time smart grid control operations. It should be noted that the eventual goal of this thesis is to develop a smart grid control system that integrates the principles of adaptive multi-agent systems and reinforcement learning, specifically using combinatorial multi-armed bandit learning. However, a learning-based adaptive multi-agent system differs significantly from a classical learning-based multi-agent system. In a classical learning-based multi-agent system, each agent determines its estimated optimal policy based on observed rewards from the environment. In contrast, in a learning-based adaptive multi-agent system, each agent seeks its estimated optimal policy based on the observed criticalities from its local environment, i.e., neighborhood. To ensure clarity for readers from diverse backgrounds, we first present the design of a classical learning-based multi-agent system in this chapter to optimize the studied smart grid problem. This system follows the well-established framework and terminologies commonly used in the literature related to the design of learning-based multi-agent systems. Building upon the system proposed in this chapter, the final learning-based adaptive multi-agent system will be introduced in the subsequent chapter. These two distinct designs will also assist readers in distinguishing between a learning-based adaptive multi-agent system and a classical learning-based multi-agent system.

This chapter begins with a discussion of the studied smart grid problem and its mathematical formulation in Section 4.1. An introduction to the multi-armed bandit problem is provided in Section 4.3, laying the groundwork for a comprehensive discussion of the proposed combinatorial multi-armed bandit-based multi-agent system in Section 4.4. Finally, the conclusion of this chapter and motivation for the next chapter are discussed in Section 4.5.

4.1 Studied smart grid problem

Multi-armed bandit is still a developing field, and the applicability of multi-armed bandit for smart grid energy management has not been studied amply in the literature. Thus, a relatively simpler (yet complex) smart grid optimization problem is studied in this chapter compared to the grid balancing problem studied in the last two chapters. Rather than directly applying multi-armed bandit to control electric vehicles (EVs) for providing energy imbalance ancillary services (studied in Chapters 2 & 3), the problem of smart charging of EVs is studied in this chapter. Hence, the basic idea is the same i.e., control EV charging optimally. But, the objective function of each EV is different. Here, the objective of each EV is to minimize its daily charging cost in the presence of uncertain PV energy generation shared among all electric vehicles, and variable electricity pricing.

Description

Smart charging can be defined as the process of intelligently controlling EV charging to optimize its total energy consumption. In particular, EVs present in the distribution grid can be utilized to maintain the grid's stability [103]. An aggregator can accumulate the existing flexibility in the distribution grid and offer it to system operators and other

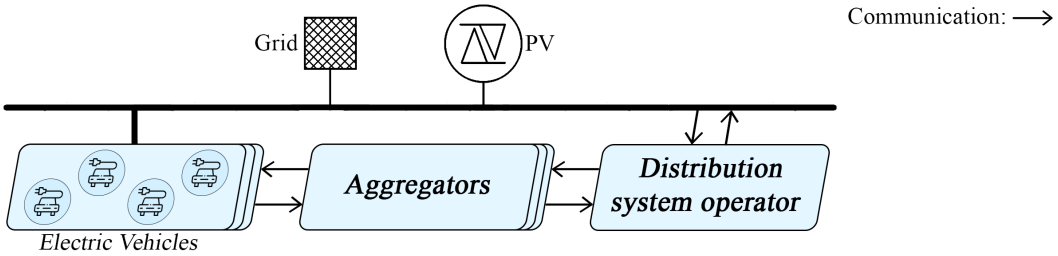

Figure 4.1: Distribution grid operation with electric vehicles and aggregators.

electricity market agents, as shown in Figure 4.1. Aggregators can encourage EVs to reduce their load demand during peak hours, and shift their charging to lower demand periods. This control would ensure the distribution network's stable operation [119].

The difference between smart charging and uncontrolled charging of EVs is depicted in Figure 4.2. The uncontrolled charging represents a worst-case scenario here as it assumes that EVs charge at their rated powers as soon as their owners come back home in the evening. It can be seen in Figure 4.2 that peak load demand occurs in the case of an uncontrolled EV charging strategy. This would cause congestion in the

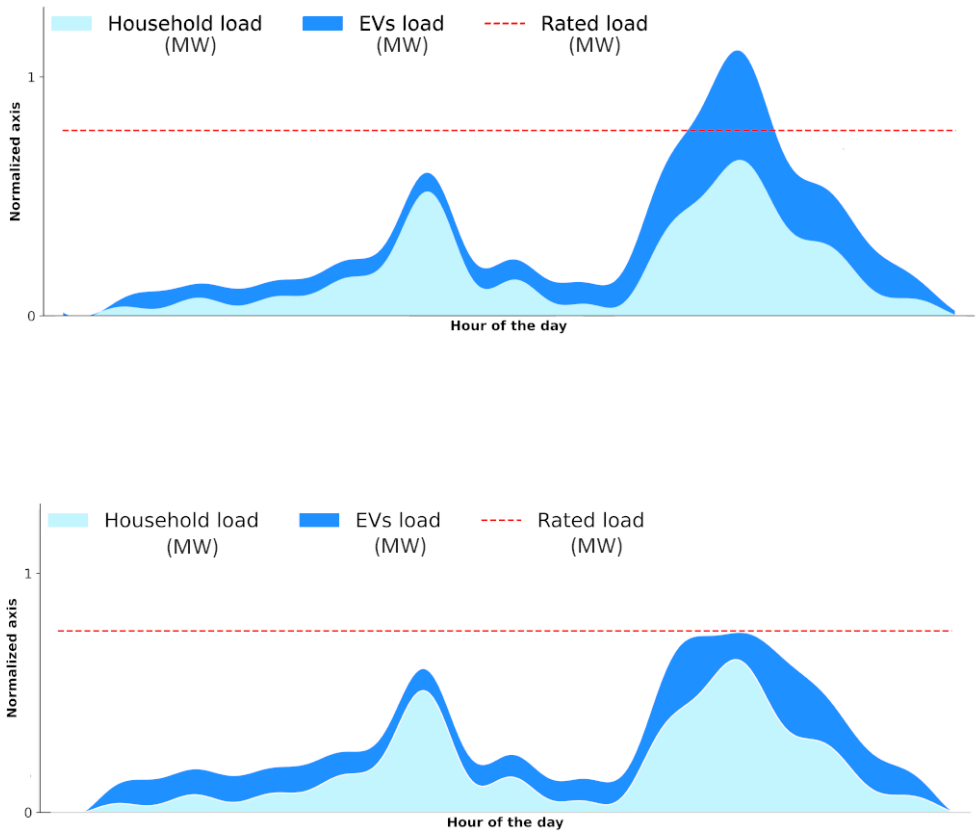

Figure 4.2: Difference between uncontrolled charging (top figure) and smart charging (bottom figure) of EVs. These curves are obtained based on the electrical network described in Section 5.3.

system. However, this peak demand may be avoided if a smart charging approach is followed [99].

A relatively simpler way to encourage EVs is through *dynamic electricity pricing* [96]. The stress on the electrical grid can be reflected through different price levels throughout the day. A higher load demand would correspond to a higher offered electricity price during a particular period. Thus, each EV would be encouraged to shift its charging hours to reduce its daily cost. This demand shift would lower stress on the distribution grid during peak hours. This mechanism establishes a fair trade between DSOs and EVs, with each aggregator acting as an intermediate party. However, this more straightforward price-based strategy has drawbacks in a decentralized environment.

Dynamic price-based charging strategy

This is a rule-based charging strategy. The objective of each EV is to minimize its daily charging cost. This can be performed easily by following Algorithm 4.1. The algorithm executes in three steps:

- The algorithm starts with each EV receiving the daily dynamic electricity price profile $c(t)$.
- Then, each EV selects its best-charging instants I (charging instants with the lowest prices). The number of selected instants $||I||_1$ should be equal to the number of charging instants required to achieve the desired state of charge I_{req}, assuming charging at EV's rated power.
- Finally, EV e charges at its rated power $P_{e,a,max}$ during the selected instants.

Algorithm 4.1 Dynamic price-based charging (each EV)

Require: Received daily dynamic electricity pricing signal for each instant $c(t)$
Require: Total charging instants required to achieve the desired SoC C_{req}
Require: Rated charging power of EV $P_{e,a,max}$
1: $t_{arrive} \coloneqq$ EV's arrival time
2: $t_{depart} \coloneqq$ EV's departure time
3: Find $I = \arg\min_t c(t)$ s.t. $||I||_1 = C_{req}$ & $I \in [t_{arrive}, t_{depart})$
4: Charge during the selected I instants

Indeed, this algorithm may work (satisfy all prosumers and DSOs) if not all EVs receive the same dynamic electricity price signal [155]. However, to ensure fairness among different EVs, the responsible aggregator must gather additional information from EVs (such as their arrival times and departure times). This approach would form a centralized model, which would suffer from the inherent drawbacks of centralization [3]. In a decentralized setting, when each EV is observing the same dynamic electricity price signal and trying to minimize its daily charging cost by executing Algorithm 4.1, the desired grid stability is not guaranteed. It is because all EVs would start charging during lower electricity price periods, which would simply shift the peak load demand

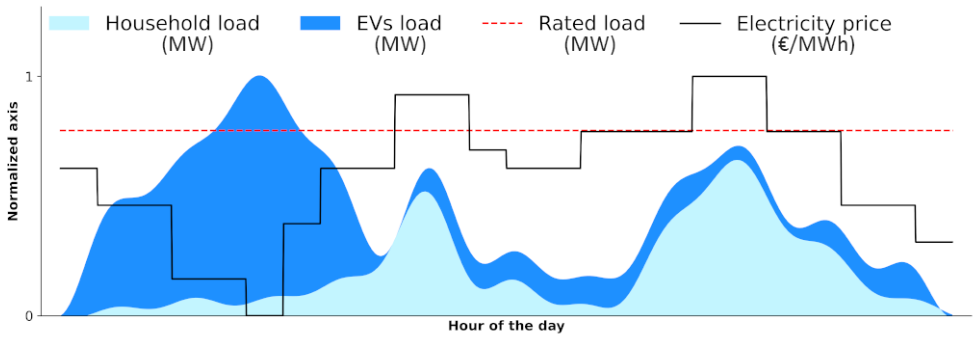

Figure 4.3: Dynamic price-based charging of EVs. These curves are obtained based on the electrical network described in Section 5.3.

to new instants during the day rather than eliminating it. This phenomenon is known as the avalanche effect [45], and it is visible in Figure 4.3. Thus, compared to Algorithm 4.1, a more sophisticated decentralized algorithm is required.

Impact of PVs

The presence of photovoltaics (PVs) in modern electrical grids increases the degree of challenges [146]. The penetration of PVs is growing in the form of modern photovoltaic power stations [165]. Electric vehicles can utilize the electrical energy generated from these PV stations without any monetary cost [25], [180], [195]. This could reduce the stress on the grid and the operation of fossil fuel power stations. However, instantaneous energy generation from PVs depends on various technological and environmental factors; thus, it is intermittent [73]. This intermittency increases the complexity of the studied smart charging problem, as uncertainty in PV energy production can not only increase the total daily charging costs of EVs, but also affect the grid's stability (in case of significant forecasting errors) [101]. Thus, EVs must learn the uncertainty in this freely offered PV energy. In the following sub-section, the studied smart charging problem is described mathematically.

Mathematical formulation

The studied smart charging problem can be descriptively summarized as follows:

- **Objective:** The objective is to minimize the daily charging cost of each EV in the distribution grid. Indeed, this can be achieved by charging EVs at cheaper electricity prices instant during the day. Furthermore, it is assumed in this study that each EV can utilize the energy generated by PV power stations without paying any cost. Thus, this freely available PV energy should be used to the maximum extent.

- **Constraints:** A set of constraints must be satisfied in the studied smart charging problem. First of all, the network should remain stable at all times to meet the DSO constraints. These constraints include no electrical current congestion,

and no voltage limit violation. Secondly, the constraints of prosumers must be satisfied as well. These constraints include sufficient charging of each EV before its departure. Some physical constraints (i.e., general load flow constraints, and constraints related to the state of each EV's battery) must also be followed.

Objective function modeling

There are E EVs in the studied distribution grid. Each day d, each EV makes the decision of charging (or not charging) from the grid at each instant t. Each instant $t \in [0, n]$ is associated with an instantaneous electricity price $c(t)$. The charging power of EV e, connected to bus a, at instant t, is represented by the variable $P_{e,a}(t) \in [0, P_{e,a,max}]$, where $P_{e,a,max}$ is the rated charging power of EV e, which is connected to bus a. This variable (i.e., $P_{e,a}(t)$) is the decision variable of the studied optimization problem. Furthermore, Δt is the resolution (duration of each decision interval i.e., second, minute, hour, etc.) of the optimization problem. The daily charging cost of EV e, connected to bus a, is given by $C_{e,a}(d)$ on the d-th day. The described objective function can be written mathematically as stated in Equation (4.1):

Smart charging problem's objective function

$$\min_{P_{e,a}(t)} \sum_{e=1}^{E} C_{e,a}(d) = \min_{P_{e,a}(t)} \sum_{e=1}^{E} \sum_{t=1}^{n} c(t) P_{e,a}(t) \Delta t - \sum_{e'=1}^{E} \sum_{e=e'+1}^{E} |C_{e',a,pu}(d) - C_{e,a,pu}(d)| \tag{4.1}$$

The right-hand side of the given objective function comprises two terms. The first term depicts the total daily charging cost paid by each EV. The second term takes the fairness among EVs into account, by making sure that the differences among per-unit charging costs (cost per energy unit) of all EVs are minimized. The per-unit charging cost $C_{e,a,pu}(d)$ of EV e for day d is calculated as:

$$C_{e,a,pu}(d) = \frac{\sum_{t=1}^{n} c(t) P_{e,a}(t) \Delta t}{\sum_{t=1}^{n} P_{e,a}(t) \Delta t}. \tag{4.2}$$

It should be emphasized that incorporating this additional term to address fairness within the objective function is but one of several available options, including lexicographic optimization, Pareto optimization, and multi-objective optimization, among others. Furthermore, the described minimization of differences among per-unit charging costs of all EVs is formulated as a soft constraint, and not as a hard constraint in this smart charging problem. This is because per-unit charging costs depend on the arrival and departure times of EVs. It is possible that an EV may not be connected to the grid during cheaper electricity price instants, and thus its per-unit charging cost may not be as low as other EVs, that were present in the network during cheaper electricity price instants.

Constraints' modeling

The set of hard constraints comprises the physical constraints of the distribution network, the constraints of DSOs, and the constraints of prosumers. It is required that the power flows in any electrical distribution must obey Ohm's law (i.e., *voltage* = *current* $*$ *resistance*) [67]. These constraints are stated in Equations (4.3)-(4.7):

Distribution network's physical constraints

$$P_a(t) = P_{a,gen}(t) - P_{a,dem}(t) \tag{4.3}$$

$$Q_a(t) = Q_{a,gen}(t) - Q_{a,dem}(t) \tag{4.4}$$

$$P_{ab}(t) + iQ_{ab}(t) = V_a(t)\left(V_a^*(t) - V_b^*(t)\right)Y_{ab}^* \tag{4.5}$$

$$\sum_b P_{ab}(t) = P_a(t) \tag{4.6}$$

$$\sum_b Q_{ab}(t) = Q_a(t) \tag{4.7}$$

Instantaneous active power $P_a(t)$ at bus a is equal to the difference between total generated $P_{a,gen}(t)$ and total demanded power $P_{a,dem}(t)$. A similar equation can be modeled for instantaneous reactive powers $Q_a(t)$ at each bus a. Equation (4.5) relates voltages at bus a and bus b (i.e., $V_a(t)$ and $V_b(t)$ respectively) with admittance matrix Y_{ab}^* of electrical line between bus a and bus b. Equations (4.6) and (4.7) make sure that the inflow of powers is equal to the outflow of powers at each bus.

In a smart grid, distribution system operators (DSOs) are responsible for keeping the grid stable by following its set of constraints [107], [153]. These constraints are given in Equations (4.8)-(4.11):

Distribution network operator's constraints

$$I_{ab}(t) \leq I_{ab,max} \tag{4.8}$$

$$V_{a,min} \leq |V_a(t)| \leq V_{a,max} \tag{4.9}$$

$$P_a(t) \leq P_{a,max}(t) \tag{4.10}$$

$$Q_a(t) \leq Q_{a,max}(t) \tag{4.11}$$

There should not be any congestion in a smart grid, as well as voltage magnitudes should remain within a suitable range. The congestion constraint is given by Equation (4.8), which states that the root-mean-square current flowing through the electrical line between bus a and bus b should not be greater than its rated value. The magnitude of the instantaneous root-mean-square voltage at each bus $V_a(t)$ is also bounded between a maximum value $V_{a,max}$, and a minimum value $V_{a,min}$, Equation (4.9). Distribution

system operators (DSOs) or energy providers must also limit the instantaneous power drawn at each bus a in the network, Equations (4.10) & (4.11).

The state of charge (SoC) of a battery is the level of charge of an electric battery relative to its capacity. It is generally expressed as a percentage. The satisfaction of each prosumer is a constraint as well in this optimization problem. Each electric vehicle (EV) e, connected to bus a, should have a pre-defined minimum state of charge $SoC_{e,adepart}$ at its departure time $t_{e,a,depart}$. This constraint would allow each EV owner to have a smooth journey. Additionally, to slow down battery degradation, there are bounds placed on the state of charge (SoC) of each EV [55]. The SoC of each EV $SoC_{e,a}(t)$ should be between a set maximum SoC value $SoC_{e,amax}$, and a set minimum SoC value $SoC_{e,amin}$. The instantaneous SoC value of a battery depends on its past instant's SoC value, its capacity $E_{e,a,bat}$, its charging/discharging efficiency $\eta_{e,a}$, and its instantaneous charging power $P_{e,a}(t)$. The state of health (SoH) of an electric vehicle $SoH_{e,a}(t)$ should also be a positive integer. The SoH variable of the battery aids in determining how much it has degraded over time. Its range of values is 0 to 1. In this formulation, the end-of-life of a battery is defined as a capacity fade of 20%. An EV battery must be replaced when it reaches the end of its useful life. Term $E_{e,atp}$ stands for the energy throughput of an EV battery. It is defined as the total amount of energy that batteries can store and discharge during their lifetime. The throughput of a battery depends on its capacity, its efficiency, its cycle life, and its depth of discharge. These prosumer constraints are listed as follows in Equations (4.12)-(4.14):

Prosumer's constraints

$$SoC_{e,amin} \leq SoC_{e,a}(t) = SoC_{e,a}(t-1) + \frac{P_{e,a}(t)\eta_{e,a}\Delta t}{E_{e,a,bat}} \leq SoC_{e,amax} \tag{4.12}$$

$$SoH_{e,a}(t) = SoH_{e,a}(t-1) - \frac{P_{e,a}(t)\Delta t}{0.2E_{e,a,tp}} > 0 \tag{4.13}$$

$$SoC_{e,a}(t_{e,a,depart}) \geq SoC_{e,adepart} \tag{4.14}$$

Now that the targeted smart charging problem has been defined, different optimization techniques can be applied to solve the stated problem. In the upcoming subsection, the concepts of multi-armed bandit are introduced.

4.2 Relevant research and scope

Related work

Electric vehicle control solutions have been extensively studied in recent years to provide support to the grid. A centralized control strategy to minimize the total charging costs of electric vehicles is presented in [194]. Constraints of prosumers have been considered in the mentioned study. However, no DSO constraints were considered. In [127], the voltage stability constraint of the DSO has been considered. However, limits on the electrical current in network lines have not been considered. Charging

cost minimization of EVs has been achieved using stochastic mixed integer linear programming in [149]. However, the architecture of the proposed solution is centralized. The control solutions in [194], [127], and [149] may lack scalability in real-life due to their centralized solution architecture.

In [100], a hierarchical MAS has been proposed to find optimal charging strategies for electric vehicles. No distributed energy resources were considered in this study. A hierarchical MAS is presented in [81] to control the EV charging while avoiding congestion in the system. Hierarchical MAS based on heuristic control has also been developed in [75] to minimize the cost of supporting a micro-grid using EVs. However, no DSO constraints were considered in this study. In [142], a hierarchical agent-based control system to coordinate the charging of EVs has been presented. A hierarchical MAS based on quadratic optimization has also been studied in [126] to incorporate demand response and coordinated charging of EVs in distribution networks. However, these hierarchical solutions in [100], [81], [75], [142], and [126] may still suffer from inherent drawbacks of centralization to some extent.

An internet-inspired scalable MAS has been proposed in [175] to optimize the charging of EVs. However, the proposed system requires an accurate model of the distribution system for its functioning. For a number of proposed decentralized MAS solutions to work, an accurate distribution system model must be available. These necessary models are frequently inaccurate or completely unknown, which makes it difficult or impossible to use these methods in a real-world situation. To tackle this issue, a model-free adaptive MAS based on reinforcement learning has been presented in [207] for optimal charging of electric vehicles to maintain the grid's stability. An adaptive decentralized MAS has been developed in [57] by defining simplistic actions for each EV agent. However, no DSO constraints have been considered in their study. Multi-armed bandit-based systems have also been proposed to control the charging of EVs in [117] and [199]. However, the architectures of both these systems are not fully decentralized.

Scope

The scope of the studied smart charging problem is defined in Figure 4.4. The contributions and limitations of the proposed solution become more apparent through the proper definition of its scope.

More focus is given to architecture, scale, control's temporal resolution, considered constraints, stochasticity, and comparison aspects in this dissertation, as shown in Figure 4.4. To tackle the potential drawbacks of centralization, the *architecture* of the proposed system is modeled to be decentralized. The designed decentralized system should be able to optimize smart grids of large *scale* (>10,000 EVs). Additionally, it should be able to perform this *control* in real-time to tackle instantaneous uncertainties arising from DERs. The developed decentralized system should also consider grid *constraints* and prosumer constraints while performing optimization. Furthermore, full *stochasticity* (and not pseudo-stochasticity) is considered in daily PV energy production. In this studied problem, the *comparison* between the proposed system and other baseline strategies is made both in terms of optimality and scalability.

A comparatively lower focus is given to the DERs, evaluation, charging technol-

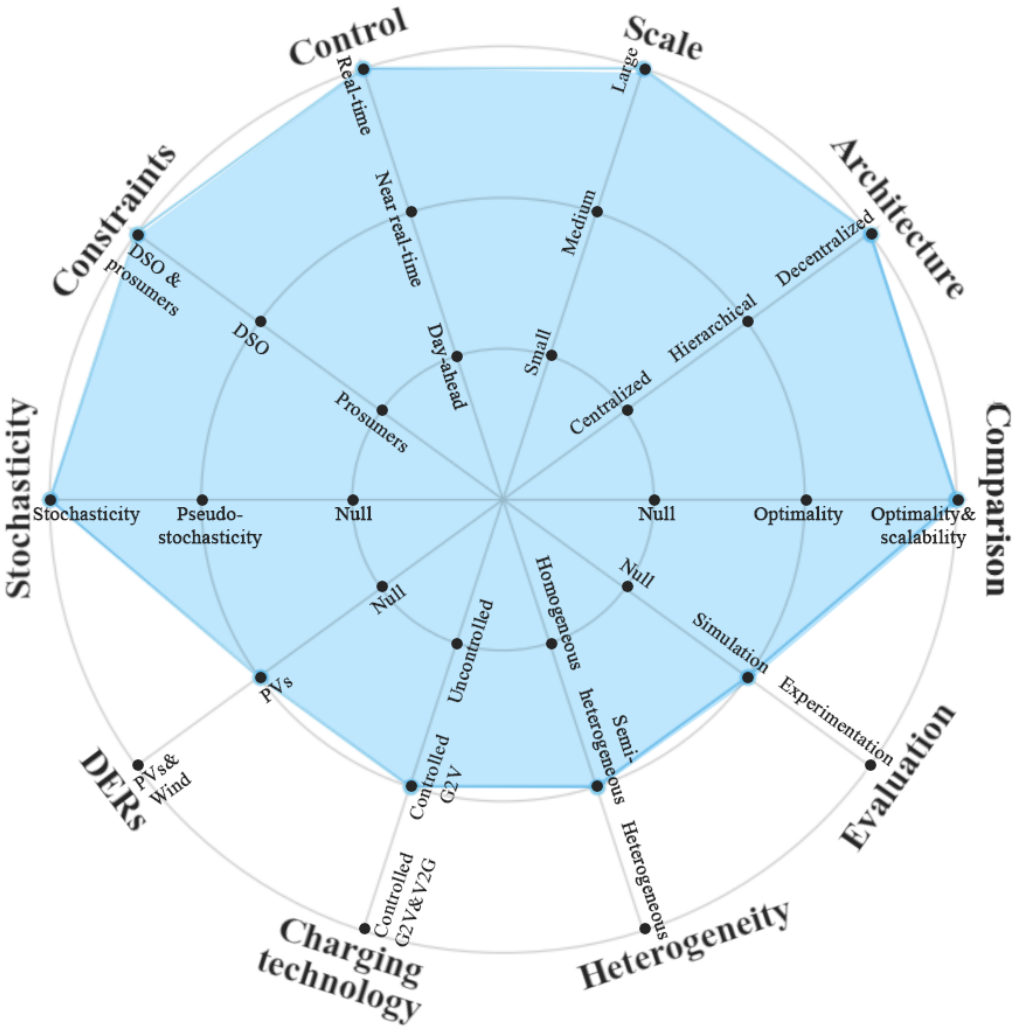

Figure 4.4: Scope of the studied smart grid problem in this chapter.

ogy, and heterogeneity aspects in this study, as shown in Figure 4.4. Only the penetration of PVs as *DERs* is considered at the distribution level in this studied problem. For *evaluation*, simulation case studies are performed. Only controlled grid-to-vehicle (G2V) *charging technology* is considered here. It is a first-stage assumption as this study is the first to propose a fully decentralized smart charging system that uses the concepts of combinatorial multi-armed bandit, to the best of the author's knowledge. Thus, the goal here is also to evaluate the optimality of just using controlled G2V technology and the need for controlled V2G technology for decentralized smart charging. *Heterogeneity* is related to the diversity of electric vehicles present in a smart grid. Hybrid EV models are considered in the simulation case studies here. Hybrid means that EVs have heterogeneous characteristics (initial SoC, arrival time, departure time, and battery capacity) as well as some homogeneous characteristics (charging efficiency, desired final SoC, and rated charging power). However, the developed system is expected to work efficiently even if all EVs are fully heterogeneous.

4.3 Introduction to multi-armed bandit

The multi-armed bandit problem is a subclass of reinforcement learning, that comes under artificial intelligence's umbrella. *Artificial intelligence* (AI) can be defined as any technique that would enable computers to mimic human intelligence [6]. *Machine learning* (ML) is a subset of AI, that utilizes data or statistics-based algorithms to improve the system's performance over time, by replicating how humans learn. It can be further divided into three main categories [151]:

- Supervised machine learning
- Unsupervised machine learning
- Reinforcement machine learning

Supervised machine learning

Supervised learning is a process of inputting properly labeled data, along with the correct outputs, to train the machine learning algorithm for proper classification of the data, or to predict an outcome. The learning model adjusts its parameters, based on the input data. The training process is finished when the fitting is complete i.e., no significant change in the parameters is observed. This type of process can be used to group a large quantity of data into different sets (classification), as well as to predict an outcome by understanding the relationship between different variables (regression) [132]. Some of the most commonly used supervised learning algorithms include k-nearest neighbor, naïve Bayes, linear regression, logistic regression, random forest, decision trees, and support vector machine (SVM) [34]. These learning algorithms have found extensive applications in real-life problems such as, customer retention [157], spam electronic-mail classification [152], weather forecasting [202], and fraud detection [48].

Unsupervised machine learning

Unsupervised learning algorithms are used to learn hidden patterns by analyzing and clustering unlabelled input data. Unlike supervised learning, the input data is not properly labeled, and does not consist of correctly mapped outputs. The learning algorithm itself finds the differences and similarities among the input data. Commonly used unsupervised machine learning algorithms include k-means clustering, association rules, and probabilistic clustering [17]. These algorithms are being used in different practical applications including data analysis [22], pattern recognition [33], image recognition [191], and customer segmentation [173].

Reinforcement machine learning

Reinforcement machine learning algorithms are designed to imitate the learning abilities of humans through experiences. Similar to a child learning the difference between a bad action and a good action, and hence navigating through life with the goal of making the most amount of good actions; a reinforcement learning algorithm learns to

differentiate between a good action and a bad action through its interactions with its environment [166]. Every good action results in a positive reward for the learning algorithm, and the goal is to maximize the running sum of these observed rewards. These algorithms are now being deployed to tackle learning tasks such as, in self-driving cars [104], in the navigation of robots [206], in playing games [185], or in making real-time decisions [201].

Note 4.3.1

Reinforcement learning algorithms are beneficial especially when sufficient data is unavailable to reasonably carry out supervised or unsupervised learning (i.e., a perfect oracle is unknown), which is generally the case in smart grid control.

Reinforcement learning can be further divided into the following categories [161]:

- Standard reinforcement learning
- Multi-armed bandit learning

To understand the differences between standard reinforcement learning and multi-armed bandit learning, some essential elements of any reinforcement learning problem are defined below:

Reinforcement learning definitions

Definition 4.3.1 (Agent)**.** An agent is a computer program that performs different tasks continuously and autonomously, on behalf of humans.

Definition 4.3.2 (Environment)**.** The environment of an agent consists of everything that surrounds the agent and with which it interacts.

Definition 4.3.3 (Reinforcement learning agent)**.** A reinforcement learning agent is a computer program that interacts with the environment through its actions, and it receives rewards based on its actions.

Definition 4.3.4 (State)**.** The state of a reinforcement learning problem is defined as all the observations that the learning agent receives from its environment at a given time.

Definition 4.3.5 (Action)**.** The action of a reinforcement learning agent is the mean through which it can interact with its environment.

Definition 4.3.6 (Reward)**.** The reward in a reinforcement learning problem is the signal that a learning agent receives from its environment based on its action at a given time.

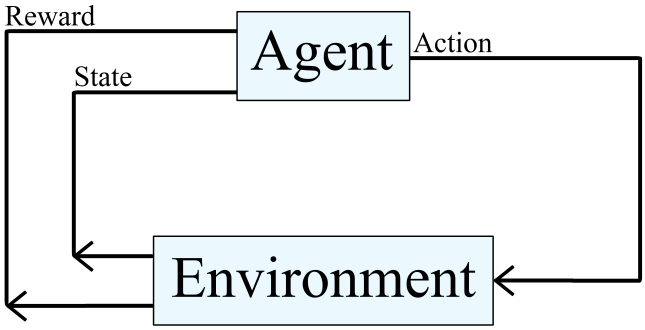

Figure 4.5: Agent-environment interaction in standard reinforcement learning.

In both types of reinforcement learning, the basic idea is the same i.e., the learning agent interacts with its environment through actions, and the goal of the agent is to maximize its cumulative reward. However, the main difference exists in the framework of both types. In standard reinforcement learning, the problem is formulated as a Markov decision process (a problem that satisfies the Markov property) [166]. The Markov property is the memory-less property for stochastic processes, which can be simply defined as: *the future is only dependent on the present and not the past*. In standard reinforcement learning, it is assumed that the action made by the agent can transition the existing state of the environment. Thus, the learning agent holds the information regarding the current state of the environment, the action made by the agent, the observed reward, and the next state of the environment, as shown in Figure 4.5. This enables the agent to handle more complex problems such as, playing games [185], navigation in real-world [104], and making complex decisions in real-time [201]. The most commonly used standard reinforcement learning algorithms include Q-learning, SARSA, policy gradient, deep Q-learning, and actor-critic algorithms [166].

In contrast, multi-armed bandit (MAB) learning is a simpler sub-class of reinforcement learning [161]. The goal of a multi-armed bandit agent is the same i.e., to maximize its cumulative reward by finding the best possible action(s) through interactions with its environment. However, in the simplest multi-armed bandit setting, it is assumed that an environment holds only a single state, and an agent's action does not change the existing state of its environment. This is shown in Figure 4.6. Hence, multi-armed bandit algorithms belong to a simpler subset of the Markov decision process, resulting in a better theoretical understanding and convergence guarantees of existing MAB algorithms than the current standard RL algorithms with function approximations.

k-armed bandit problem

The *multi-armed bandit* problem, also called the k-armed bandit problem, consists of a learning agent repeatedly selecting an action a, from a set $[A]$ of available k number of actions. At a given time, the selected action generates a reward $r \in \mathbb{R}$, which is sampled from a stationary probability distribution associated with the selected action. The goal of a learning agent is to maximize its cumulative reward by playing the best available action. This is termed as *exploitation*. In the described setting, a learning

Note 4.3.2

It is important to keep in mind that, despite the fact that the most simplistic multi-armed bandit scenario is shown here with only one environment state. It is possible to model more advanced multi-armed bandit settings (i.e., switching bandits, adversarial bandits, and non-stationary stochastic bandits) in which the environment is allowed to have state transitions [70], [12], and [5]. Therefore, it's important to dispel the misconception that MAB can only have one state.

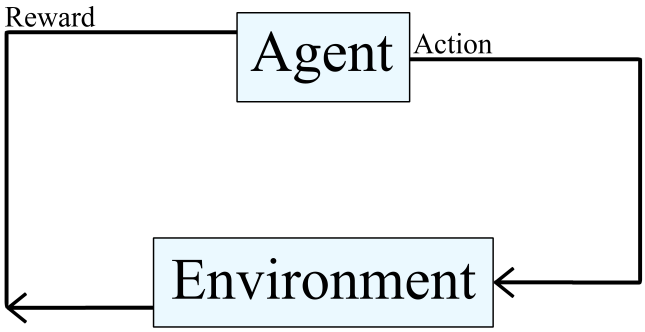

Figure 4.6: Agent-environment interaction in multi-armed bandit learning.

algorithm observes the reward for its selected action only, and not for the actions that were not selected. Therefore, a learning algorithm needs to go through an *exploration* phase. During this exploration phase, a learning algorithm may try different actions to learn the expected outcome (reward) of selecting each action. Thus, this learning algorithm is faced by the *exploration-exploitation* dilemma: to continuously explore by selecting different actions, but to exploit the already learned information as well by selecting the expected best action [166]. The name, multi-armed bandit, comes from the scenario faced by a gambler [183], shown in Figure 4.7. In Figure 4.7, there are k slot machines available. Each slot machine follows a reward distribution unknown to the player (gambler). Our player (gambler) can play any available slot machine and receive a reward value. Thus, the goal of our player (gambler) is to find, as soon as possible, the slot machine with the highest estimated reward value.

Note 4.3.3

Generally, multi-armed bandit learning algorithms are expected to exhibit faster convergence rates in comparison to other widely used reinforcement learning algorithms that employ function approximations, like deep Q-learning. In practical smart grid control applications, where access to a perfect oracle is limited, and agents must interact with their environment in an online manner to estimate the best policy, this faster convergence becomes highly advantageous as it directly translates to increased economic benefits.

Multi-armed bandit algorithms can be sub-categorized depending on the choice

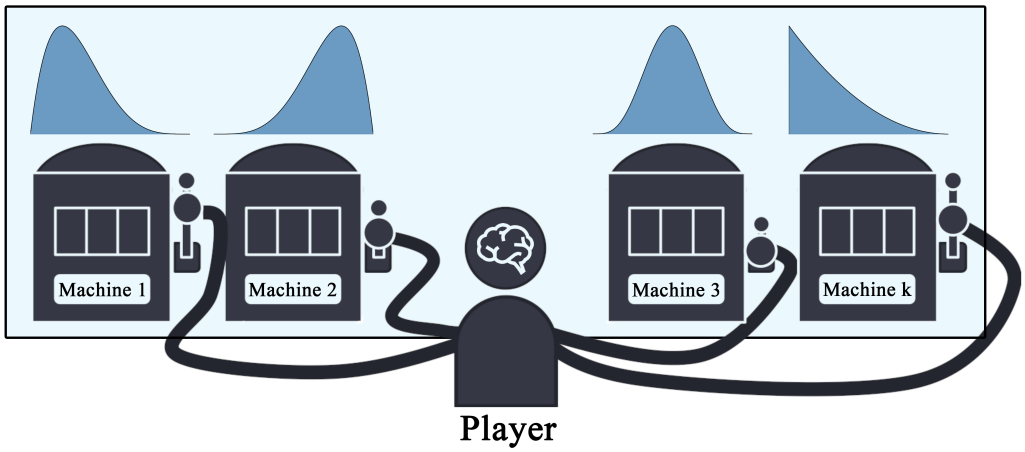

Figure 4.7: Multi-armed bandit framework: exploration vs. exploitation dilemma.

of different design variables, such as the type of feedback from the environment, the reward model, and the availability of any contextual information:

- **Feedback:** The feedback observed by a learning algorithm can be divided into three main types: *bandit feedback*, when only the reward for the selected action is observed by our learning algorithm; *full feedback*, when the rewards for all available actions are observed; *semi-bandit feedback*, when some extra information along with the reward for the selected action is revealed [161].

- **Reward model:** The reward (in *stochastic bandit*) is considered to be sampled from a stationary probability distribution depending only on the selected action i.e., *i.i.d (independent and identically distributed)* reward. The reward (in *adversarial bandit*) can also come from an adversary trying to minimize the reward of our learning algorithm i.e., *adversarial reward* [12].

- **Contextual information:** There may be additional information available beforehand that can assist a MAB learning algorithm in picking the best possible action i.e., *contextual bandit* [120]. This approach is in contrast to the situations when no extra information is available to improve the decision-making in bandit i.e., *non-contextual bandit*.

Multi-armed bandit algorithms have found practical applications in different domains. They can be used in training a robot to manage a variety of tasks such as, advertisement [114], recommendation system [31], Internet communication [44], [205], [121], and smart grid optimization [94], [38]. To better understand the functionality of different existing multi-armed bandit algorithms, some critical definitions are required.

Multi-armed bandit definitions

Definition 4.3.7 (Expected reward)**.** In stochastic bandit with i.i.d. rewards, the reward is considered to be sampled from a stationary probability distribution. Thus, the expected reward $Q(a)$ of selection action a is defined as:

$$q(a) = \mathbb{E}[r|a] \tag{4.15}$$

Definition 4.3.8 (Learning objective)**.** The training in bandit is generally online. During each round of play $t = 1, 2, 3, ..., T$, the learning algorithm selects an action, and observes the corresponding reward $r(t)$ for that round. The learning objective of the bandit algorithm is to maximize the cumulative reward i.e., $\sum_{t=1}^{T} r(t)$.

Definition 4.3.9 (Optimal action)**.** The optimal action a^* is defined as the action that would generate the maximum expected reward. It can be mathematically written as:

$$a^* = \arg\max_{a \in A} q(a) = \arg\max_{a \in A} \mathbb{E}[r|a] \tag{4.16}$$

Hence, the learning objective of a MAB algorithm is to find this optimal action a^*.

Definition 4.3.10 (Optimal expected reward)**.** The optimal (best) expected reward for selecting the action a^* during one round of play is given as:

$$q^* = \max_{a \in A} q(a) = \max_{a \in A} \mathbb{E}[r|a] \tag{4.17}$$

Definition 4.3.11 (Optimal policy)**.** The optimal policy π for any multi-armed bandit problem is the policy that maximizes the learning agent's cumulative reward during T rounds of play.

Multi-armed bandit definitions (cont'd)

Definition 4.3.12 (Expected pseudo-regret)**.** The performance of any multi-armed bandit algorithm can be evaluated based on the pseudo-regret of the algorithm. The expected pseudo-regret $R(T)$ is defined as the difference between the cumulative expected reward q^* of always selecting the optimal action a^*, and the cumulative expected reward during T rounds of the algorithm's learning. The term $a(t)$ represents the action selected during round t. The expected pseudo-regret can be written as:

$$\begin{aligned}\mathbb{E}[R(T)] &= \max_{a \in A} \mathbb{E}\left[\sum_{t=1}^{T} q(a) - \sum_{t=1}^{T} q(a(t))\right] \\ &= Tq^* - \mathbb{E}[\sum_{t=1}^{T} q(a(t))]\end{aligned} \tag{4.18}$$

Generally, sub-linear regrets are expected from good multi-armed bandit algorithms i.e., $\mathbb{E}[R(T)]/T \longrightarrow 0$. The expected pseudo-regret will be null when the optima action a^* is selected during all T rounds of play by the learning algorithm. The theoretical pseudo-regret lower bound of MAB algorithms is given in [109].

Definition 4.3.13 (Expected precision)**.** Another interesting metric to benchmark the performances of different multi-armed bandit algorithms is the expected precision of the algorithm. The expected precision is the expected number of times, the optimal action a^* is selected. It is defined as:

$$\mathbb{E}[P(T)] = \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\{a(t) = a^*\}\right] \tag{4.19}$$

The precision value ranges between 0 (optimal action is never selected) and 1 (optimal action is always selected).

Frequentist bandit learning

There exists a wide variety of multi-armed bandit algorithms. The choice of a bandit algorithm to tackle a problem depends on the problem itself, its constraints, and available additional information. Frequentist bandit learning is a popular learning approach. In frequentist learning, it is assumed that the repetitive sampling from a population would help in finding the single true value of the parameter governing that population [134]. Some of the most commonly used frequentist bandit algorithms are explained as follows. In later sections, these algorithms will serve as the base for developing more complex bandit algorithms for smart grid optimization.

Uniform exploration

This algorithm works on the philosophy of *first exploration, then exploitation* [161]. Uniform exploration bandit learning algorithm will try each action a, from the set of all available actions $[A]$, a fixed N number of times at the beginning of its learning phase. After each action has been tried N number of times, our learning algorithm will select the action with the highest expected reward in its remaining playing rounds. The functionality of the uniform exploration algorithm is summarized in Algorithm 4.2. The algorithm goes through three main stages i.e, exploration phase, finding the estimated optimal action, and the exploitation phase. Terms $N_t(a)$, and $Q_t(a)$ represent the number of times action a has been selected, and the estimated expected reward of action a, at the end of round t. During the exploration phase, the algorithm selects each action N number of times. The reward $r(t)$ is observed for the selected action a, in round t. Afterward, the estimated reward of the selected action is updated based on the observed reward. The estimated reward is calculated by simply taking the average of all observed rewards:

$$\begin{aligned} Q_t(a) &= \frac{r(1) + r(2) + r(3) + ... + r(N_t(a))}{N_t(a)} \\ &= \frac{1}{N_t(a)} \sum_{t=1}^{N_t(a)} r(t) \end{aligned} \tag{4.20}$$

The drawback of this simplistic calculation in Equation 4.20 is that the history of past observed rewards is required. This means that the memory and computational requirements are unbounded and will increase over time. However, this can be easily handled through *incremental updates*: update $Q_t(a)$ after each round, and then use this updated $Q_t(a)$ in the next round to calculate $Q_{(t+1)}(a)$ [166]. It is, in fact, derived using the equation 4.20:

$$\begin{aligned} Q_t(a) &= \frac{1}{N_t(a)} \sum_{t=1}^{N_t(a)} r(t) \\ &= \frac{1}{N_t(a)} \left(r(N_t(a)) + \sum_{t=1}^{N_t(a)-1} r(t) \right) \\ &= Q_{(t-1)}(a) + \frac{1}{N_t(a)} \left(r(t) - Q_t(a) \right) \end{aligned} \tag{4.21}$$

Note 4.3.4

The uniform exploration multi-armed bandit algorithm holds regret bound $\mathbb{E}[R(T)] \leq T^{2/3} O(k \log T)^{(1/3)}$, when $N = T^{2/3} O(k \log T)^{(1/3)}$ [161]. The O(.) stands for the big O notation which studies the limiting behaviors of functions/algorithms.

Algorithm 4.2 Uniform exploration

Require: Total number of arms (actions) k
Require: Number of times each action should be selected during exploration N

$N(a) := 0$
$Q(a) := 0$
▷ exploration phase
for $t = 1, 2, 3, ..., (k * N)$ **do**
    for $a = 1, 2, 3, ..., (k)$ **do**
        if $N(a) < N$ **then**
            Select action a
            Observe reward $r(t)$
            $N(a) := N(a) + 1$
            $Q(a) := Q(a) + \frac{1}{N(a)}(r(t) - Q(a))$
        end if
    end for
end for
▷ Finding the estimated optimal action
Find the estimated optimal action $\hat{a}^* := \arg\max_a Q(a)$ (arbitrary tie braking)
▷ exploitation phase
for $t = (k * N) + 1, (k * N) + 2, (k * N) + 3, ..., T$ **do**
    Select action $\hat{a}^*$
end for

Epsilon-greedy

The main drawback of the *uniform exploration* algorithm is that the length of the exploration phase is critical in determining the *expected precision* of the algorithm. A fairly large value of N could decrease the cumulative reward of the learning algorithm. This can be handled using the *epsilon-greedy* approach [166]. In computer sciences, a greedy approach is an approach that selects the best short term option. Hence, instead of forcing exploration only at the beginning of the learning, the algorithm is encouraged to explore, with a fixed (but low) probability ε, throughout its duration of operation. The *ε-greedy* algorithm is summarized in Algorithm 4.3. Initialization of the ε-greedy algorithm is similar to the uniform exploration algorithm (Algorithm 4.2), but with an additional parameter ε. At the beginning of each round, the ε-greedy algorithm samples a random value from a uniform distribution $\mathcal{U}(0, 1)$. If this sampled random number is lower than the input ε, then an action is selected at random. Otherwise, the action with the highest expected reward value is selected. At the end of each round, a reward is observed for the selected action, and $Q(a)$ is updated.

Note 4.3.5

The presented epsilon-greedy multi-armed bandit algorithm achieves regret bound $\mathbb{E}[R(t)] \leq t^{2/3} O(k \log t)^{(1/3)}$, when $\varepsilon = t^{-1/3}(K \log t)^{1/3}$ [161].

Algorithm 4.3 Epsilon-greedy

Require: Total number of arms (actions) k
Require: Exploration probability ε
$N(a) := 0$
$Q(a) := 0$
for $t = 1, 2, 3, ...$ **do**
    $\alpha \sim \mathcal{U}(0, 1)$
    if $\alpha < \varepsilon$ **then**
▷ exploration phase
        $a(t) :=$ Select a random action a
    else
▷ exploitation phase
        $a(t) :=$ Select the action $\arg\max_a Q(a)$
    end if
    Observe reward $r(t)$
    $N(a(t)) := N(a(t)) + 1$
    $Q(a(t)) := Q(a(t)) + \frac{1}{N(a(t))}(r(t) - Q(a(t)))$
end for

Epsilon-decay

It is evident that the uniform exploration (Algorithm 4.2), and the ε-greedy algorithm (Algorithm 4.3) do not achieve a sub-linear regret. This is due to continuous exploration throughout the lifespan of this learning algorithm. This problem can be solved by decreasing the ε-greedy algorithm's learning rate as time progresses [166]. This is achieved by decreasing the ε value with time. The functioning of the ε-decay algorithm is presented in Algorithm 4.4. Compared to Algorithm 4.3, the main difference is that the ε value is dependent on the time here. With time, estimations of our learning algorithm regarding the return of each arm are expected to become more solid. Hence, the exploration is desired to decrease. Here, the exploration is controlled using the learning parameter c. This parameter is directly proportional to the degree of exploration of the presented ε-decay algorithm, Algorithm 4.4. If a higher value of c is given, then the algorithm will perform more exploration, but the exploration cost of the algorithm will increase as well. On the other hand, the exploration cost can be reduced through a lower value of c, but this will reduce the degree of exploration as well.

Upper confidence bound

Upper confidence bound (UCB) is the most well-known multi-armed bandit algorithm [11] . The philosophy behind this algorithm is: *optimism in the face of uncertainty*. In UCB, the optimal action is selected according to the following rule, derived using Hoeffding's inequality [79]:

$$a(t) = \arg\max_{a \in A} \left(Q(a) + c\sqrt{\frac{\log t}{N(a)}} \right) \tag{4.22}$$

Algorithm 4.4 Epsilon-decay

Require: Total number of arms (actions) k
Require: Exploration parameter c
$N(a) := 0$
$Q(a) := 0$
for $t = 1, 2, 3, ...$ **do**
    $\varepsilon = \frac{c}{c+t}$
    $\alpha \sim \mathcal{U}(0, 1)$
    if $\alpha < \varepsilon$ **then**
▷ exploration phase
        $a(t) :=$ Select a random action a
    else
▷ exploitation phase
        $a(t) :=$ Select the action $\arg\max_a Q(a)$
    end if
    Observe reward $r(t)$
    $N(a(t)) := N(a(t)) + 1$
    $Q(a(t)) := Q(a(t)) + \frac{1}{N(a(t))}(r(t) - Q(a(t)))$
end for

The first part of Equation 4.22 encourages *exploitation* by utilizing the already obtained knowledge of the environment. The action with the highest estimated return will be selected, if only this part is followed. The second part of Equation 4.22 encourages *exploration*. The value of this second term is directly related to the number of times a specific action has been selected in the past. A low precision rate of an action would correspond to a lower $N(a)$ value, which would add a larger uncertainty in the said action. Hence, the selection of this action will become more likely. As time progresses, the denominator of this term will increase. As a result, this term will start shrinking, and the algorithm will become more focused on the exploitation part. The illustration

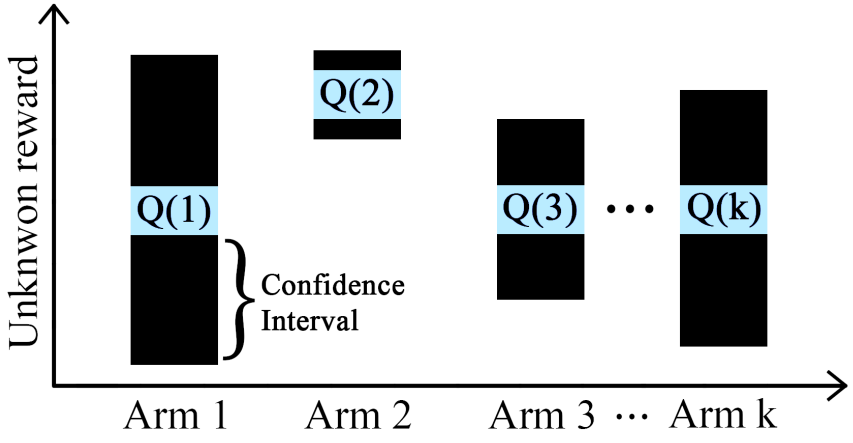

Figure 4.8: Illustration of the UCB principle.

of this UCB approach is shown in Figure 4.8. The UCB algorithm has estimated mean rewards $Q(a)$ for playing each arm a. The confidence interval shows the amount of uncertainty in $Q(a)$. The pseudo-code of UCB is presented in Algorithm 4.5.

Algorithm 4.5 Upper confidence bound

Require: Total number of arms (actions) k
Require: Exploration parameter c
$N(a) :=$ Number of times action a is selected
$N(a) \longleftarrow 0$
$Q(a) :=$ Estimate of the expected reward of action a
$r(t) :=$ Observed reward in round t
for $t = 1, 2, 3, ...$ **do**
    $a(t) :=$ Select the action $\arg\max_a \left(Q(a) + c\sqrt{\frac{\log t}{N(a)}} \right)$
    Observe reward $r(t)$
    $N(a(t)) \longleftarrow N(a(t)) + 1$
    $Q(a(t)) \longleftarrow Q(a(t)) + \frac{1}{N(a(t))}(r(t) - Q(a(t)))$
end for

The degree of exploration is controlled using the *exploration parameter* i.e., c. Similar to the ε-decay strategy (Algorithm 4.4), a higher value of c would result in more exploration at the expense of a higher exploration cost. This exploration cost can be controlled by reducing the value of c, but that would also discourage the explorative nature of the algorithm. The UCB algorithm achieves a sub-linear regret.

Note 4.3.6

The upper confidence bound algorithm's regret, after T rounds of play, is bounded by $\mathbb{E}[R(T)] \leq const. \left(\frac{k \log T}{\Delta}\right)$, where Δ is the distance between a sub-optimal arm and the optimal arm [13].

Exponential-weight algorithm for exploration and exploitation

The exponential-weight algorithm for exploration and exploitation (Exp3) is a famous adversarial bandit algorithm [12]. In adversarial bandit algorithms, it is considered that the learning algorithm is competing against an adversary. The return for each action is no longer sampled from a stationary probability distribution, instead the payoff structure for each action is selected by an adversary. Evidently, the philosophies followed by the UCB class of bandit algorithms and that of the adversarial bandit algorithms are quite opposite in nature. In UCB, *optimism* under uncertainty is encouraged. However, in adversarial bandit, the learning algorithm is modeled to compete against an adversary. This gives birth to a more *pessimistic* sub-class of bandit algorithms. It can be argued that payoffs in most real-life problems are not entirely adversarial. Thus, this sub-class of bandit algorithms may perform better than their theoretical guarantees, when applied to real-life applications. Nonetheless, adversarial bandit help model

competitive practical payoff scenarios e.g., trading, multi-player competitive games etc.

In the Exp3 algorithm, a set of weights $w(a)$, one weight for each available action a is maintained. The probability of an action getting selected $p(a)$ is dependent on these weights. These weights are increased (or decreased) based on the observed reward $r(t)$. To incorporate *exploration*, the algorithm selects an action at random (depending on their weights) with probability $1-\gamma$, and it performs uniform random exploration with probability γ. Its working is presented in Algorithm 4.6.

Algorithm 4.6 Exponential-weight algorithm for exploration and exploitation

Require: Total number of arms (actions) k
Require: Exploration parameter γ
$w(a) :=$ Weight for each action a
$w(a) \longleftarrow 1$
$r(t) :=$ Observed reward in round t
for $t = 1, 2, 3, ...$ **do**
    for $a = 1, 2, 3, ..., k$ **do**
        $p(a) := (1-\gamma)\frac{w(a)}{\sum_{i=1}^{k} w(i)} + \gamma\frac{1}{k}$
    end for
    $a(t) :=$ Select an action randomly based on the probabilities $p(a)$
    Observe reward $r(t)$
    for $a = 1, 2, 3, ..., k$ **do**
        if $a = a(t)$ **then**
            $\hat{r}(a) := r(t)/p(a)$
        else
            $\hat{r}(a) := 0$
        end if
        $w(a) \longleftarrow w(a)\exp(\gamma\hat{r}(a)/k)$
    end for
end for

Note 4.3.7

The Exp3 algorithm also achieves a sub-linear regret, and is bounded by $\mathbb{E}[R(T)] \leq O\left(\sqrt{kT\log(k)}\right)$ [12].

Bayesian bandit learning

Contrary to the frequentist approach, in Bayesian learning the learning parameter is not represented by a signal true value, rather it is represented by a probability distribution. This probability distribution captures the uncertainty present in the studied unknown variable. Thompson sampling is the most commonly used Bayesian learning algorithm.

Thompson sampling

Thompson sampling (TS) is a natural randomized Bayesian algorithm to tackle the exploration-exploitation dilemma in MAB problems. It was proposed initially by Thompson in 1933 [171]. However, the application of Thompson sampling for reinforcement learning was proposed in 2000 [164]. In Thompson sampling, the history $\mathcal{H}$ of past observations (action and reward pairs i.e., $(a(t), r(t))$) is modeled by a parametric likelihood function $P(r|a,\theta)$, parameterized by θ. This unknown parameter θ includes the uncertainty in the expected reward for any specific action. This uncertainty is represented using a probability distribution $P(\theta)$, known as the *prior* distribution. This probability distribution is the initial belief of the learning algorithm regarding the unknown parameter θ. Once the algorithm observes new information (history $\mathcal{H}$ is updated), the algorithm updates its initial belief regarding the unknown parameter θ using the Bayes rule [97]:

$$P(\theta|\mathcal{H}) \propto P(\mathcal{H}|\theta)P(\theta) \tag{4.23}$$

This updated belief is known as the *posterior* distribution. After the update, the learning algorithm samples the $\tilde{\theta}$ from the posterior distribution, and uses it to pick the expected best action. The objective in each round is to maximize the observed reward, based on the played action, which depends on the posterior belief i.e., $\mathbf{E}\left[r(t)|a(t),\tilde{\theta}\right]$.

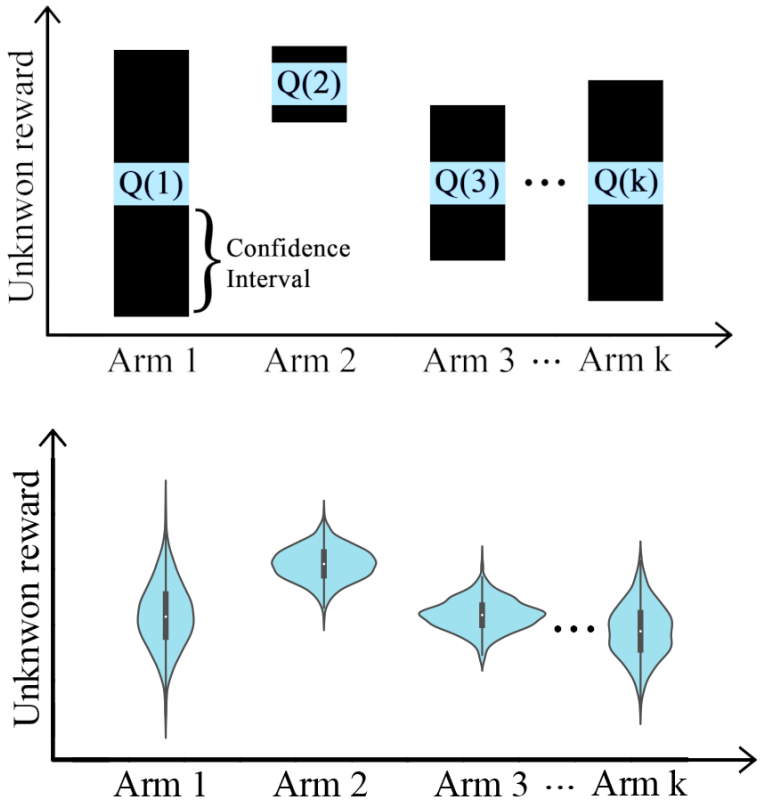

Figure 4.9: Illustration of the differences between UCB (top) and Thompson sampling (bottom) methodologies.

In each round, the action $a(t)$ is selected with the probability:

$$\int_{\theta} \mathbb{1}\{\mathbf{E}[r(t)|a(t),\theta] = \max_{a'} \mathbf{E}[r(t)|a',\theta]\} P(\theta|\mathcal{H}) d\theta \tag{4.24}$$

The key difference between the UCB algorithm and Thompson sampling is shown in Figure 4.9. In the UCB approach, there is an upper confidence bound on our empirical estimates of each unknown variable. Whereas, there exists a distribution in Thompson sampling to estimate the mean reward of each arm.

The choices of both prior and posterior distributions depend on the studied problem. Gaussian, beta, and uniform distributions are among the most common choices [188]. Thompson sampling algorithm with Gaussian prior and posterior distributions is given in Algorithm 4.7. In the case of using Gaussian priors, the Gaussian posterior for each action is calculated using Bayes law:

$$P(\theta|\mathcal{H}) = P(\theta|(a(t),r(t))) \propto e^{-\frac{N(a(t))+2}{2}\left(\theta - \frac{\hat{\mu}(a(t))N(a(t))+r(t)}{N(a(t))+2}\right)^2} \tag{4.25}$$

where $\hat{\mu}$ is the current estimate of the likelihood of getting the maximum reward by playing action $a(t)$. The number of samples to form an estimate $\tilde{\theta}(a)$ is usually equal to one. By taking only a single sample, exploration is promoted, as our estimate will be noisy. On the other hand, the distribution is going to be peaked as the algorithm gets more data (the variance term $\frac{1}{N(a)+1}$ will decrease). Thus, our single sample estimate will improve as the distribution's variance decreases.

Algorithm 4.7 Thompson sampling

Require: Total number of arms (actions) k

$N(a) := 0$
$\hat{\mu}(a) = 0$
for $t = 1, 2, 3, ...$ **do**
    for $a = 1, 2, 3, ..., k$ **do**
        $\tilde{\theta}(a) \sim \mathcal{N}(\hat{\mu}(a), \frac{1}{N(a)+1})$
    end for
    $a(t) :=$ Select the action $\arg\max_a \tilde{\theta}(a)$
    Observe reward $r(t)$
    $N(a(t)) := N(a(t)) + 1$
    $\hat{\mu}(a(t)) := \frac{\hat{\mu}(a(t))N(a(t))+r(t)}{N(a(t))+2}$
end for

Note 4.3.8

The Thompson sampling algorithm with k arms, has the expected sub-linear regret $\mathbb{E}[R(T)] \leq O\left(\sqrt{kT\ln(k)}\right)$ [1].

Combinatorial multi-armed bandit

Combinatorial multi-armed bandit (CMAB) has been proposed in the literature to handle combinatorial optimization problems. A *Combinatorial optimization problem* consists in finding the optimal object, among a finite set of objects, that satisfies a set of constraints [141]. Some of the most common combinatorial optimization problems are route planning [122], task scheduling [42], and knapsack problem [158]. In fact, the studied problem of smart charging in electrical grids also falls under the umbrella of combinatorial optimization problems. These problems may be solved using classical mathematical optimization approaches such as, linear programming [23]. However, these algorithms belong to the *NP* (nondeterministic polynomial time) class of optimization problems. Thus the efficiency of such algorithms, in terms of the time taken to obtain the optimal solution, is still an open question in theoretical computer science. The traveling salesman problem is a famous NP-hard combinatorial optimization problem [7]. It states that: *"Given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city exactly once and returns to the origin city?"* The complexity of this problem depends greatly on the number of available cities n_{cities}, as the number of possible routes is equal to $\frac{n_{cities}(n_{cities}-1)}{2}$. This complexity is illustrated in Figure 4.10. Each node in this figure represents a city, whereas each edge depicts a route connecting two cities. It can be observed that the density of these graphs increases drastically as the total number of cities are increased, thus making solving a large-scale combinatorial optimization problem a complex task.

Note 4.3.9

The *NP* (nondeterministic polynomial time) class of optimization problems are the problems that can be solved by a nondeterministic Turing machine in polynomial time. However, can a deterministic Turing machine solve these problems in polynomial time is still an unanswered question, i.e., $P = NP$? [16]

Compared to brute-force methodologies, a combinatorial multi-armed bandit algorithm can tackle combinatorial optimization problems by utilizing the feedback obtained from its environment [37]. In combinatorial multi-armed bandit (CMAB), a learning algorithm selects n actions from a set of m possible actions, corresponding to $\binom{m}{n}$ possibilities. There are m number of total available actions, also known as *base arms*. The set $[m] = \{1, 2, 3, ..., m\}$ is the set of available m arms (actions).

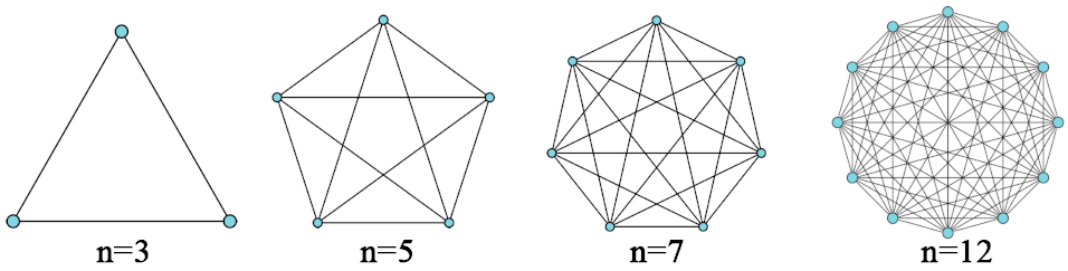

Figure 4.10: Illustration of the traveling salesman problem. Here $n_{cities} = n$.

Each action in $[m]$ is associated with a random variable $X_i(t)$, where $1 \leq i \leq m$ and $t \geq 1$. Variable $X_i(t)$ corresponds to the random outcome of the i-th arm, in its t-th round of learning. This set of random variable $[X(t)] = \{X_1(t), X_2(t), X_3(t), ..., X_m(t)\}$ is assumed to be independent and identically distributed according to some unknown set of distributions $[\mathcal{D}] = \{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, ..., \mathcal{D}_m\}$ with the set of unknown expectations $[\mu] = \{\mu_1, \mu_2, \mu_3, ..., \mu_m\}$.

In CMAB, our learning algorithm selects a subset of base arms, which is called a *super arm*, $S(t) \in \mathcal{S}$. The super arm selected by a CMAB algorithm in round t is $S(t)$. There is also a constraint $\mathcal{S} \subseteq 2^{[m]}$, where $2^{[m]}$ is the set of all possible subsets of arms. The expected reward $r_\mu(S(t))$ of playing $S(t)$ depends on the selected super arm itself and on the expectation vector of all arms $[\mu]$. In each round, a CMAB learning algorithm picks one super arm. Then, its environment provides the feedback (rewards) to our CMAB learning algorithm. This feedback depends on selected $n \in S(t)$ base arms in a given round, as well as on the type of feedback i.e., bandit feedback, semi-bandit feedback, or full feedback. The observed reward can be an aggregated reward based on the played super arm i.e., *bandit feedback*. A learning algorithm can also observe rewards for each selected base arm i.e., *semi-bandit feedback*, or it can also observe the outcomes of all possible actions (base arms) i.e., *full feedback*. Based on this information the learning objective, the optimal super arm, and the pseudo-regret can be defined as follows:

Combinatorial multi-armed bandit definitions

Definition 4.3.14 (Learning objective)**.** The learning objective of a combinatorial multi-armed bandit algorithm is to maximize its expected cumulative reward i.e., $\sum_{t=1}^{T} r_\mu(S(t))$.

Definition 4.3.15 (Optimal policy)**.** The optimal policy π is a policy that, when followed, gives the maximum expected cumulative reward. The goal of a learning agent is to find this policy.

Definition 4.3.16 (Optimal super arm)**.** The optimal (best) super arm S^* is the combination of base arms that maximizes the expected reward. It is defined as:

$$S^* = \arg\max_{S \in \mathcal{S}} \mathbb{E}[r_\mu(S)] \tag{4.26}$$

Definition 4.3.17 (Pseudo-regret)**.** The goal of a CMAB algorithm is to minimize its pseudo-regret:

$$\begin{aligned} \mathbb{E}[R(T)] &= \sum_{t=1}^{T} \mathbb{E}\left[\max_{S \in \mathcal{S}} r_\mu(S)\right] - \sum_{t=1}^{T} \mathbb{E}\left[r_\mu(S(t))\right] \\ &= T\,\mathbb{E}\left[r_\mu(S^*)\right] - \sum_{t=1}^{T} \mathbb{E}\left[r_\mu(S(t))\right] \end{aligned} \tag{4.27}$$

In this CMAB formulation, if the number of selected base arms is equal to one ($n =$

1), then it becomes the classical MAB problem. However, the time to evaluate all arms may increase exponentially for larger values of total arms m. Also, the information regarding the outcomes of underlying arms can be shared by different super arms, which is not the case in the classical MAB framework [37]. The CMAB framework allows any learning algorithm to select a collection of actions in each round, instead of selecting just one action in each round. It is worth mentioning that the presented CMAB formulation is not a specific bandit algorithm, but instead it is a framework that can be integrated with other learning strategies. Thus, the bound on the pseudo-regret of any CMAB algorithm depends on the choice of strategy that has been used to learn the set of unknown random variables $[X(t)]$. The generalized CMAB framework is presented in Algorithm 4.8.

Algorithm 4.8 Generalized combinatorial multi-armed bandit

Require: Total number of base arms m
Require: Required number of base arms in a super arm n
Require: Parameters corresponding to selected *exploration* − *exploitation* strategy
1: $S(t) :=$ Selected super arm in round t
2: $r_\mu(S(t)) :=$ Reward for play $S(t)$ super arm in round t
3: **for** $t = 1, 2, 3, ...$ **do**
4: Select $S(t) = \arg\max_{S \in \mathcal{S}} \mathbb{E}[r(S)]$ s.t. the number of base arms in $S(t) = n$
5: Observe $r_\mu(S(t))$
6: Update parameters of the selected learning strategy
7: **end for**

Regret bounds of CMAB combined with different learning strategies are given in Table 4.1. Terms n_{max} and Δ_{min} stand for the maximum number of base arms in a super arm and the minimum gap between the expected reward of the optimal solution and any non-optimal solution, respectively. It can be seen that these algorithms also achieve sub-linear regrets.

Learning strategy	Regret
UCB [37]	$O\left(\frac{n_{max}m}{\Delta_{min}}\log(T)\right)$
Exp3 [36]	$O\left(\sqrt{m^3 nT\log(\frac{n_{max}}{m})}\right)$
Thompson sampling [188]	$O\left(\frac{m}{\Delta_{min}}\log(n_{max})\log(T)\right)$

Table 4.1: Regret upper bounds for various learning strategies combined with the presented CMAB formulation.

4.4 Proposed decentralized multi-armed bandit system

In this section, the proposed smart charging system, which integrates concepts of both multi-agent systems and multi-armed bandit, is presented. At first, the mapping process of physical elements in a distribution network to software agents in a multi-armed

bandit learning framework is described. Moving forward, the communication framework which is followed by the proposed system is highlighted. Finally, CMAB algorithms are defined through which EV agents can charge smartly under uncertainties.

Multi-agent system mapping

The significant distribution grid elements, essential to optimize smart charging in the proposed system, are modeled as agents in the computerized multi-agent system. These elements are as follows:

- **Electrical line:** Each electrical line present in the distribution grid is modeled as a line agent. The goal of each line agent is to make sure that the electrical current flowing through the electrical line is below its rated value. Line agent can encourage this condition by controlling the reward that it communicates to each EV in the distribution network.
- **Electrical Bus:** Similar to electrical lines, all electrical buses of the distribution network are modeled as bus agents. Bus agents are designed to maintain the instantaneous bus voltages withing a specified threshold. Again, this can be encouraged in the network through the reward given by a bus agent to each EV.
- **Photovoltaic sensor:** A photovoltaic sensor that records the information of instantaneous energy generation is modeled as a PV agent. This PV agent has only one simple goal, which is to communicate the energy production values to each EV. Through these values, EVs can learn the PV production trend and that would help in minimizing their daily charging costs from the electrical grid.
- **Electric vehicle:** Each electric vehicle (EV) is modeled as an EV agent. Reinforcement learning abilities are implemented in each of these EV agents, which makes each of them the most important part of the whole system. These EV agents are continuously learning from the environment to minimize their daily charging costs, while satisfying the set of constraints described in the previous section.

The suggested mapping model applied to an example distribution network is shown in Figure 4.11.

Environment modeling

In a simple reinforcement learning system, a learning agent communicates (interactions through actions and rewards) with the environment [166]. This methodology also applies to the proposed smart charging system. Each EV agent is a learning agent in the system. These EV agents interact with the environment, which consists of line agents, bus agents, and PV agents. This model is also depicted in Figure 4.12. Each agent type has its own objective, which can be encouraged by each agent type through their instantaneous reward values to EV agents. In the proposed system, this communicated reward can have three possible values:

- **-1:** In case an environment agent wants to encourage an EV to stop charging at a given instant.
- **1:** In case an environment agent wants to encourage an EV to start charging at a given instant.
- **0:** In case an environment agent is not concerned by an EV's charging power at a given instant (i.e., when its constraints are not violated).

The environment part of the system is dissected, and the detailed functionality of each agent type is explained as follows.

Line agent

Congestion can occur when the magnitude of electrical current flowing through an electrical line exceeds its rated value. This line congestion is highly undesirable in a smart grid, as the network's stability is compromised. A line agent is designed to ensure that congestion does not happen in its distribution network. Peak load demand, created due to the saturation of EVs charging simultaneously, can cause electrical current congestion in distribution networks. A line agent can encourage several EVs to shift their charging times to avoid congestion. A line agent can provide this encouragement as a reward value to each EV agent, as the line agent belongs to the environment of this reinforcement learning-based system.

Instantaneous line agent reward value $rew_{e,l}(t)$, given by line agent l, to EV e depends on the instantaneous line current $I_{ab}(t)$, and the rated value of line current $I_{ab,max}$. If there is no electrical current congestion in the line, the line agent generates an instantaneous reward equal to zero. However, not all EVs get the maximum reward when electrical current congestion occurs. Instead, only a certain number of EVs get the maximum reward. These are the EVs that can charge simultaneously at instant t, without causing electrical current congestion in the distribution network. All remaining

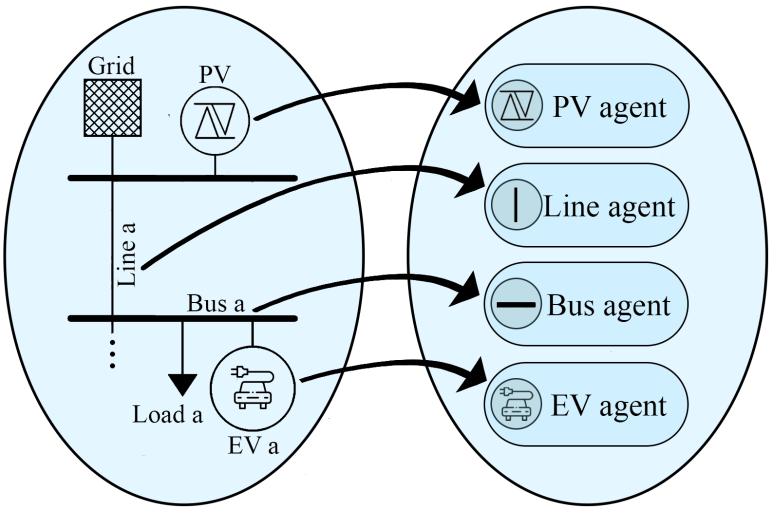

Figure 4.11: Example sub-section of a distribution network (left), and its MAS mapping (right).

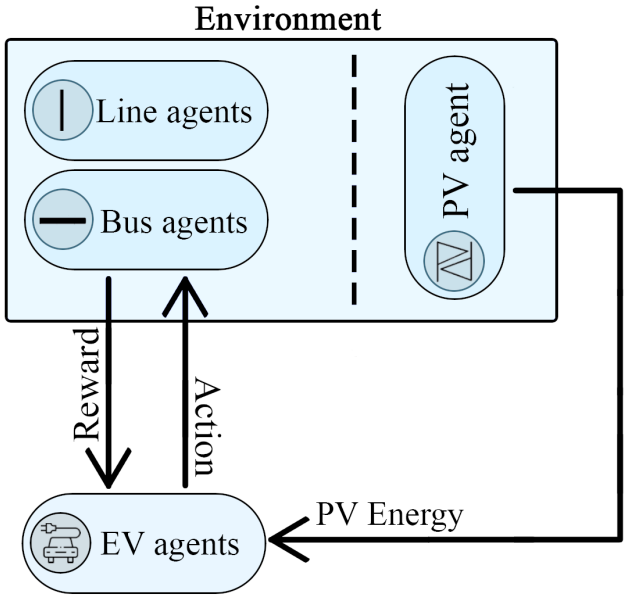

Figure 4.12: Reinforcement learning model of the proposed MAS.

EVs get the minimum possible reward value (i.e., -1). Electric vehicles (EVs) observing this lowest possible reward value would then try to find other instants to charge during the day, resulting in better reward values. The instantaneous line agent's reward value is calculated using Equation (4.28):

Line agent's reward model

$$rew_{e,l}(t) = \begin{cases} 0 & \text{if } I_{ab}(t) < I_{ab,max} \\ 0 & \text{if } I_{ab}(t) \geq I_{ab,max} \ \& \ e \in [h] \sim \mathcal{U}(0,g) \\ -1 & \text{if } I_{ab}(t) \geq I_{ab,max} \ \& \ e \notin [h] \sim \mathcal{U}(0,g) \end{cases} \tag{4.28}$$

The selection of EVs (which get the best reward value) can be modeled in different ways. One can consider EVs' departure times and each EV's remaining charging requirements to form a priority list. Then, EVs (which will observe the best reward value) can be selected on their priorities. However, such a system may not be viable for maintaining fairness among all EV agents. That is why uniform random selection (without replacement) has been utilized to select EVs randomly without any bias. Let $[g] = \{1, 2, 3, ..., g\}$ denote the set of g number of EVs charging at the same instant and causing electrical current congestion. Then, the line agent will uniformly sample $h \in [h] \sim \mathcal{U}(0, g)$ number of agents that can charge simultaneously without causing any congestion. Each EV in this uniformly sampled set of EVs $[h] = \{1, 2, 3, ..., h\}$ will receive the null reward value (i.e., 0). The remaining set of EVs $[g] - [h]$ will receive a reward value of -1. The line agent's functioning is given in Algorithm 4.9.

In Algorithm 4.9, it can be seen that each line agent also communicates $cong_l(t)$. This $cong_l(t)$ is a Boolean flag, that reflects the condition of the electrical line i.e., congested or not congested. If the electrical line is congested, then this flag is set to

Algorithm 4.9 Line agent's functionality (each line agent)

Require: Rated electrical line current $I_{ab,max}$

$t := t-$th instant of the day
$I_{ab}(t) :=$ Instantaneous electrical line current
$rew_{e,l}(t) :=$ Reward sent to EV agent e at instant t
$cong_l(t) :=$ Boolean congestion flag of line agent l at instant t
Observe $I_{ab}(t)$
Observe $[E]$
if $I_{ab}(t) < I_{ab,max}$ **then**
    $rew_{e,l}(t) := 0 \; \forall \; e \in [E]$
    $cong_l(t) :=$ False
else
    $[h] \sim U(0, g)$
    $rew_{e,l}(t) := 0 \; \forall \; [h] \cap [g]$
    $rew_{e,l}(t) := -1 \; \forall \; [g] - [h]$
    $cong_l(t) :=$ True
end if
Forward ($rew_{e,l}(t)$, $cong_l(t)$)

true, otherwise it is set to *false*. This additional variable becomes necessary as there are multiple types of "*congestion management*" agents in the environment (i.e., line agents and bus agents). Hence, line agents use this variable $cong_l(t)$ to cooperate with bus agents. This cooperation mechanism is comprehensively discussed when the bus agent's functionality and communication framework are presented next.

Bus agent

Bus agents are responsible for keeping voltage magnitudes within a desired range. Let us suppose the voltage magnitude at any electrical bus violates this condition; there will be a voltage limit violation, which is also termed as a *voltage congestion* here. Like current congestion, this voltage congestion can also lead to stability issues in a smart grid and reduce the quality of supplied electricity. Thus, a bus agent is designed to tackle this challenge. Each bus agent has to make sure that the instantaneous bus voltage $V_b(t)$ at bus b is within its specified limits i.e., $V_{b,min} < V_b(t) < V_{b,max}$. Bus agents can encourage this behavior through reward values given to EVs, similar to line agents. However, this would create two significant problems:

- If only one type of congestion management agent (i.e., line agents) were present in the system, then the system's modeling would have been relatively more straightforward. Line agents would have communicated their reward values directly to EVs. However, there are also bus agents as part of the environment as shown in Figure 4.12. Thus, both bus agents and line agents should cooperate.
- Furthermore, bus agents can encourage EVs to either charge (in case of over-voltage) or not charge (in case of under-voltage). Hence, there can be an *antagonistic scenario* in the studied system, i.e., a line agent can encourage EVs to

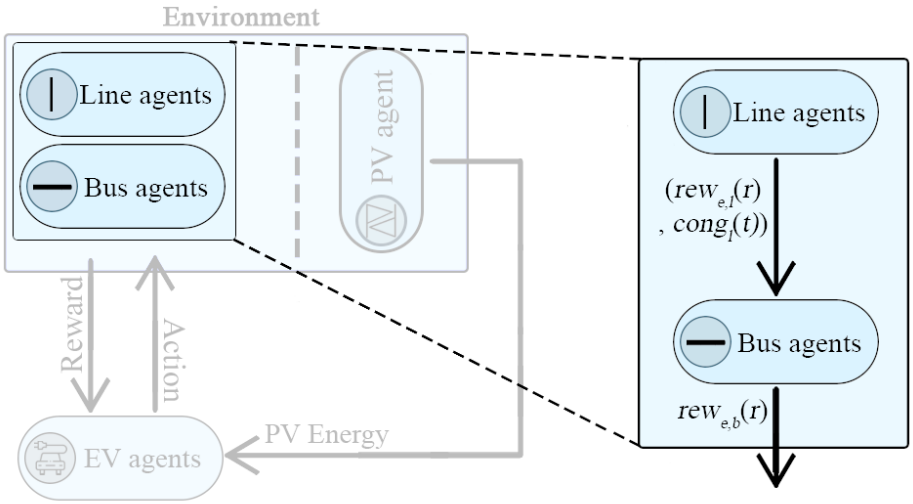

Figure 4.13: Cooperation between line agents and bus agents inside the RL environment.

stop charging (in case of electrical current congestion). In contrast, a bus agent can simultaneously encourage EVs to continue charging (in case of congestion due to bus over-voltage). This challenge further emphasizes the need for a cooperation mechanism among line and bus agents.

To incorporate cooperation, each line agent does not communicate directly with EV agents. As in a physical distribution network, each EV is electrically connected to an electrical bus. Thus, line agents communicate with EVs through bus agents. This cooperation scheme is shown in Figure 4.13.

A line agent is forwarding its instantaneous reward value $rew_{e,l}(t)$, and the state of its congestion flag $cong_l(t)$ to each bus in the system. Subsequently, each bus agent would generate the aggregated reward, which would be given to the EV agent corresponding to the EV connected to that particular bus. This solves the first problem mentioned above. To tackle the antagonistic scenario, a priority system is introduced. Voltage congestion is local congestion here, as it can be managed, at first, by controlling the connected EV's charging power. In contrast, current congestion generally impacts a more significant part of the distribution network, and managing this congestion would require controlling the charging powers of several EVs in the network. Thus, priority is given to the global congestion here, i.e., in case of simultaneous current and voltage congestions, line congestion is tackled first. Following reward model is used by bus agent b, to calculated its reward value $rew_{e,b}(t)$, for EV e, connected to bus b, at instant t:

Bus agent's reward model

$$rew_{e,b}(t) = \begin{cases} 0 & \text{if } V_{b,min} < V_b(t) < V_{b,max} \\ -1 & \text{if } V_b(t) < V_{b,min} < V_{b,max} \quad \text{(under-voltage)} \\ 1 & \text{if } V_{b,min} < V_{b,max} < V_b(t) \quad \text{(over-voltage)} \end{cases} \tag{4.29}$$

This model in Equation (4.29) has three main parts. First, if there is no voltage congestion in the system (i.e., $V_{b,min} < V_b(t) < V_{b,max}$), then the connected EV will get a null reward. In case there is an under-voltage issue (i.e., $V_b(t) < V_{b,min} < V_{b,max}$), then the connected EV will be encouraged to stop charging through a negative unity reward. Finally, if there is an over-voltage issue (i.e., $V_{b,min} < V_{b,max} < V_b(t)$), then the connected EV will be encouraged to charge through a positive unity reward. The functioning of each bus agent is described in Algorithm 4.10.

Algorithm 4.10 Bus agent's functionality (each bus agent)

Require: Maximum electrical bus voltage $V_{b,max}$
Require: Minimum electrical bus voltage $V_{b,min}$
$t := t-$th instant of the day
$V_b(t) :=$ Instantaneous electrical bus voltage
$rew_{e,b}(t) :=$ Reward sent to EV agent e at instant t
Observe $V_b(t)$
Observe ($rew_{e,l}(t)$, $cong_l(t)$) from Algorithm 4.9
if $V_{b,min} < V_b(t) < V_{b,max}$ **then**
    $rew_{e,b}(t) = 0$
end if
if $V_b(t) < V_{b,min} < V_{b,max}$ **then**
    $rew_{e,b}(t) = 1$
end if
if $V_{b,min} < V_{b,max} < V_b(t)$ **then**
    $rew_{e,b}(t) = -1$
end if
if ($rew_{e,b}(t)$ = 0 or $cong_l(t)$ = True) **then**
    Forward $rew_{e,l}(t)$ to the connected EV agent
else
    Forward $rew_{e,b}(t)$ to the connected EV agent
end if

Each bus agent starts by observing its instantaneous voltage magnitude value and by receiving a message(s) from line agent(s). Afterward, it calculates its instantaneous reward value using the presented model in Equation (4.29). Finally, it sends the calculated reward value to the connected EV's agent. This reward value depends on the received message(s). To give priority to current congestion, if a bus agent has received a request from any line agent with $cong_l(t) = True$, then it directly passes that request to the connected EV agent. Otherwise, if no line is congested, and the bus agent is congested, then it will forward its own request to the connected EV agent. This priority-based cooperation model makes sure that the proposed multi-agent multi-armed bandit system functions smoothly even under antagonistic conditions.

Photovoltaic agent

PV agent has the most straightforward role in the proposed system. It aims to transmit the instantaneous PV energy generation value to each EV agent in the system. This

Note 4.4.1

Although it is possible to manage bus voltages through reactive power control, only active power control has been considered in this study. A follow-up work should include the possibility of combining both active and reactive power control. Also, given that low-voltage distribution networks are significantly resistive, controlling active power makes as much sense as controlling reactive power. Finally, not using reactive power flows helps reduce the need for costly volt-ampere reactive (VAR) compensation mechanisms.

knowledge of PV energy production would enable EV agents to learn the trend of daily PV production. As EV agents can use the energy produced by PV power stations without any cost, thus it becomes essential for EVs to learn uncertainties in daily PV energy production, to achieve a near-optimal (ideally optimal) solution. The execution steps of each PV agent are presented in Algorithm 4.11.

Algorithm 4.11 PV agent's functionality

1: $t := t-$th instant of the day
2: $P_{PV}(t) :=$ Instantaneous PV production
3: Observe $P_{PV}(t)$
4: Forward $P_{PV}(t)$ to all EV agents

Communication framework

There are two main communication channels in the proposed smart charging system:

- First, the communication link between each learning agent and the environment.
- Second channel is the communication link within the environment. Both of these communication types are shown in Figure 4.12 and Figure 4.13.

As described in the previous subsection, both types of congestion management agents in the environment, i.e., line and bus agents, need a communication link. This communication link establishes the path through which the environment, as a whole, can communicate efficiently with an agent, leading to a smoother distribution grid operation. A line agent initiates the communication by sending an ordered pair $(rew_{e,l}(t), cong_l(t))$ to each bus agent in the system. After receiving this message, each bus agent performs its functionality as stated in Algorithm 4.10. Each EV learning agent receives two messages from the environment. Firstly, it receives the reward value generated by its bus agent $(rew_{e,b}(t))$. Secondly, it gets the instantaneous value of PV energy generation $P_{PV}(t)$ from the PV agent. Both of these messages are singleton values. Each EV learning agent uses these communicated values to optimize its daily charging cost while satisfying the desired set of constraints. The communication is also directional in the system proposed here. For example, if there is inflow line congestion then the congested line agent will only be communicating with the EVs coming downstream.

CMAB learning agent modeling

Combinatorial multi-armed bandit (CMAB) algorithms have found applications in various domains, such as optimizing advertisement selection [37], playing real-time strategy games [140], and optimizing modern communication networks [69]. Selfish CMAB multi-agent systems, with excellent practical results, have been proposed to optimize communication in Internet of things (IoT) devices [28], [24]. But, applications of multi-agent CMAB learning for smart grid control can not be found, according to the best of the author's knowledge. Electric vehicle learning agents are modeled here using the previously stated CMAB framework.

Electric vehicle agent

In the proposed MAB system, each EV is acting as a learning agent. The learning objective of each EV agent is to solve the optimization problem described in Section 4.1 by controlling its instantaneous charging. In simpler words, each EV has to select a number of instants daily, to charge from the grid, such that its total daily charging cost is minimized, while also satisfying constraints described in Section 4.1. To formulate this as a combinatorial multi-armed bandit (CMAB) problem, each day d is divided into m number of instants. The set of instants $[m] = \{1, 2, 3, ..., m\}$ represents the base arms of this CMAB formulation. Each instant $i \in [m]$ is a base arm here, and each base arm is associated with an unknown distribution $\mathcal{D}_i$ with expectation μ_i. Each base arm can be picked (or not picked) by an EV agent to charge from the grid (or to not charge from the grid). Thus, the action set $\mathcal{A}$ of each EV agent for each base arm $i \in [m]$ consists of two values only i.e., $\mathcal{A} = \{0, 1\}$. This represents the binary decision made by each EV agent regarding each base arm $i \in [m]$. If i-th base arm is selected (i.e., $= 1$), EV agent e will charge from the grid, at its rated power $P_{e,a,max}$. On the contrary, if i-th base arm is not selected (i.e., $= 0$), EV agent e will not charge from the grid at instant i during day d.

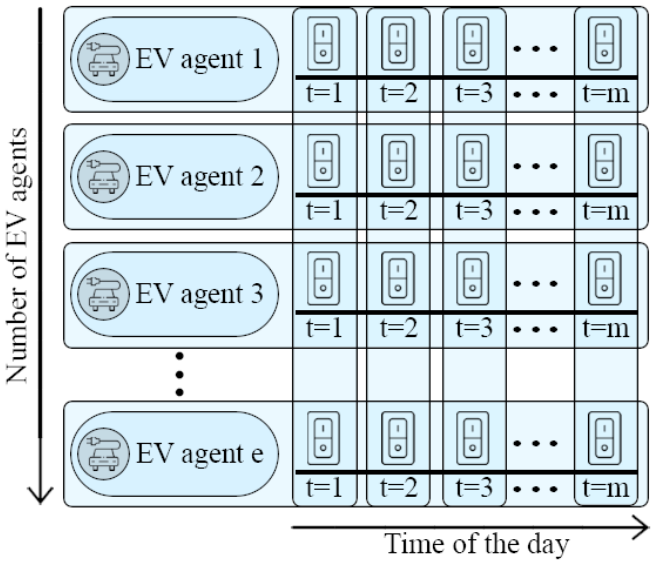

Figure 4.14: Working of the proposed decentralized CMAB algorithm.

A visual representation of this decision making process is given in Figure 4.14. Each EV has to make a binary decision of charging (at max power) or not charging from the grid during each $i \in [m]$ instant of the day. This binary decision is represented by an on-off switch in the discussed figure. Here, each EV makes decisions for itself, without doing any direct communication with other EVs. Thus, it is possible that the majority of the EV agents "turn the charging switch on" during the same instances when the electricity price is low, which may cause congestion in the system. Each EV agent utilizes the combinatorial multi-armed bandit learning framework to optimize its daily charging cost while maintaining the stability of the system. At this first stage, a binary action space has been considered in this formulation. However, this can be replaced with a discrete or continuous action space as well (i.e., the on-off switches in the discussed figure will be replaced by variable resistors allowing each EV to charge in the range $[0, P_{max}]$.

Combinatorial multi-armed bandit general assumptions

Assumption 4.4.1 (Mutually independent outcomes)**.** The outcome of each base arm in $[m]$ is according to some unknown distribution. It is assumed that these outcomes are mutually independent i.e., $\mathcal{D} = \mathcal{D}_1 \cdot \mathcal{D}_2 \cdot \mathcal{D}_3 \cdot ... \cdot \mathcal{D}_m$.

Assumption 4.4.2 (Monotonicity)**.** It is assumed that the expected reward of playing a super arm $S(t) \in \mathcal{S}$ is a monotonically non-decreasing function with respect to the expectation vector, i.e., $\forall\ i, i' \in [m]$, if $\mu_i \leq \mu_{i'}$, then $r_\mu(S(t)) \leq r_{\mu'}(S(t))$ for all $S(t) \in \mathcal{S}$.

Assumption 4.4.3 (Lipschitz-continuity)**.** There exists a constant B such that for every pair of expectations μ and μ', $|r_\mu(S(t)) - r_{\mu'}(S(t))| \leq B||\mu_{S(t)} - \mu'_{S(t)}||_1$ is satisfied for all $S(t) \in \mathcal{S}$. Here $\mu_{S(t)}$ stands for the projection of vector $[\mu]$ on $S(t)$.

As it is a combinatorial optimization problem, each EV agent would have to select a combination of base arms i.e., the EV agent will have to select a combination of instants to charge from the grid to minimize its daily charging cost. This combination of selected instants to charge from the grid is also called a *super arm* in the CMAB setting. Let $S_e(d) \in \{0,1\}^m$ indicates the super arm selected by EV agent e on day d, where $S_{e,i}(d) = 1$ shows that i-th instant, during day d, is selected by EV e, to charge from the grid. Whereas, $S_{e,i}(d) = 0$ means that instant i, during day d, has not been selected by EV e to charge from the grid. After playing super arm $S_e(d)$, EV agent e will observe rewards from its environment through *semi-bandit feedback*. The daily reward of EV agent e is defined as the sum of rewards for each selected charging instant $i \in [m]$, during day d. This daily reward depends on the price of electricity and the condition (congested or not congested) of the electrical network at all selected instants (base arms). Each EV agent tries to maximize its expected reward $r_\mu(S_e(d))$, which depends on the selected super arm by EV agent e, on day d, and on the vector of expectations of all base arms. For the studied smart charging problem, the following

two additional CMAB assumptions are made:

Combinatorial multi-armed bandit problem-specific assumptions

Assumption 4.4.4 (Equally spaced base arms)**.** If $\mathbf{D}([m]_i)$ gives the duration of i-th instant (base arm), then it is assumed that $\mathbf{D}([m]_i) = \mathbf{D}([m]_{i+1}) \; \forall \; 0 \leq i \leq m-1$, i.e., the duration of all instants are equal.

Assumption 4.4.5 (Linearly structured super arms)**.** Let $\theta_e(d) \in \mathbb{R}^m$ stand for the vector of unknown learning parameters of EV agent e, on day d. It is assumed here that super arm $S_e(d) \in \mathcal{S}$, played by each EV agent e, on day d, follows a linear structure, i.e., $\mathbb{E}\left[r_\mu(S_e(d))\right] = (S_e(d))^T \cdot (\theta_e(d))$. This assumption helps tackle the complexity of combinatorial optimization problems.

Here, each element of $\theta_e(d) \in \mathbb{R}^m$ vector is a learning parameter (an unknown random variable) associated with each base arm in $[m]$. The goal of an EV agent e is to learn this unknown vector $\theta_e(d)$. An EV agent improves its estimation based on its observed reward values after each interaction with its environment. This obtained reward from the environment can guide an EV agent towards solutions that do not cause any congestion in the distribution network. However, each EV agent also needs to minimize its daily charging cost, which can be done by charging when the electricity price is lowest during the day. Thus, before updating its estimate of the unknown parameters vector $\theta_e(d)$, an EV agent applies the following reward model:

EV learning agent's reward model

$$rew_{e,i}(S_e(d)) = \begin{cases} rew_{e,env,i}(S_e(d)) & \text{if } rew_{e,env,i}(S_e(d)) \neq 0 \\ 1 - c(i) & \text{if } rew_{e,env,i}(S_e(d)) = 0 \end{cases} \tag{4.30}$$

Here, $rew_{e,env,i}(S_e(d))$ is the reward received by EV agent e from its environment, corresponding to instant (base arm) $i \in S_e(d)$, on day d (i.e., $rew_{e,env,i}(S_e(d))$ is the output of the Algorithm 4.10). Term $rew_{e,i}(S_e(d))$ represents the final reward observed by EV agent e, for selecting base arm $i \in S_e(d)$, after playing super arm $S_e(d)$ on day d. This final EV reward is directly used to update elements of the unknown learning parameters vector. The idea is that, if an EV is receiving a non-zero value from its environment, then the distribution grid is congested. Thus, the final reward of EV agent e will be according to the received value from its environment i.e., an EV agent will be helping the distribution grid in avoiding all congestion.

On the other hand, when the environment reward (i.e., the reward representing the state of the distribution network) is null, each EV agent will calculate its final reward (and thus update the unknown learning parameters vector) based on the instantaneous normalized electricity price $c(i)$[1]. This way EV agent e will be learning to differentiate

[1]The impact of this reward function's part on the performance of the proposed system is discussed in Appendix A

between the cost of charging at each instant. Based on this information, each EV agent can select its optimal policy i.e., it would select instants (base arms) that are expected to minimize its daily charging cost (without causing any congestion in the distribution network). Optimal policy π_e of EV agent e can be defined as follows:

Linear CMAB optimal policy

Definition 4.4.1 (Optima policy π_e)**.** The optimal policy π_e can be obtained by EV agent e when θ_e is completely known. According to this optimal policy, EV agent e will play the following optimal arm S_e^*:

$$S_e^* = \arg \max_{S_e(d) \in \{0,1\}^m} (S_e(d))^T \cdot \theta_e \tag{4.31}$$

Pseudo-regret of the learning EV agent based on the described optimal policy π_e can be calculated as follows:

Linear CMAB pseudo-regret

Definition 4.4.2 (Pseudo-regret)**.** The pseudo-regret $\mathbb{E}[R_e(D)]$ observed by EV agent e, after D days of learning is given as:

$$\begin{aligned} \mathbb{E}[R_e(D)] &= \sum_{d=1}^{D} \mathbb{E}[r_\mu(S_e^*)] - \sum_{d=1}^{D} \mathbb{E}[r_\mu(S_e(d))] \\ &= D \cdot \mathbb{E}[r_\mu(S_e^*)] - \sum_{d=1}^{D} \mathbb{E}[r_\mu(S_e(d))] \end{aligned} \tag{4.32}$$

The learning goal of an EV learning agent is to minimize this regret value by learning the unknown vector $\theta_e(d)$. However, in the studied smart charging problem, learning this unknown vector $\theta_e(d)$ is not completely straightforward. Each EV has to learn this unknown parameters vector under uncertainties. There can be three major sources of uncertainties in the system here:

- **PV uncertainty:** Daily PV production is intermittent and variable [73]. It depends on several technological and environmental factors. Thus, uncertainty is involved in this daily PV production value from an EV agent's perspective. This uncertainty is crucial here as an EV can use the produced PV energy free of cost, hence optimizing its daily charging cost. But, if an EV does not have an estimation of this PV uncertainty, then the obtained solution (its daily charging cost) can be far from optimal.
- **Real-time uncertainty:** In the formulation discussed so far, EV agent e is selecting a super arm $S_e(d)$, i.e., it is picking the estimated best instants to charge

from the grid, during day d. This selection is only made at the start of day d. However, in practical life, there can be different uncertainties. Information used by EV agent e, to select $S_e(d)$ can be changed during the day e.g., the EV owner can change the desired value of $SoC_{e,depart}$. Thus, it is desirable that the designed system should be able to manage these real-time uncertainties.

- **Opponents' actions uncertainty:** The proposed system is a multi-agent system. Thus, there is also uncertainty in the choice of super arms of other agents, from one EV agent's point of view. As other agents are learning as well, this uncertainty is a non-stationary random variable. Each intelligent EV agent should also learn to perform optimization in the presence of this uncertainty.

To tackle the *PV uncertainty*, Bayesian learning is applied. Bayesian learning is a great tool to find the true value of an unknown random variable by updating our initial beliefs, based on the latest observed data [97]. This learning rule is defined in Equation (4.23). Let $\varphi_e(d) \in \mathbb{R}^m$ be the vector of instantaneous PV energy production values during m instants of day d, of EV agent e. This unknown vector has to be learned by EV agent e through Bayesian learning. Each EV agent can use this information to learn the trend of freely available PV energy production during each day, and then it can use that information to calculate the remaining required number of charging instants from the grid, to achieve its desired SoC $SoC_{e,depart}$. Let $i_{e,req}(d)$ be the number of instants required by EV e to charge from the grid, on day d. This value can be calculated by each EV as follows:

PV uncertainty management model

$$i_{e,req}(d) = \left\lceil \frac{60E_{bat}(SoC_{e,depart} - SoC_{e,ini})}{||[m]_i||_1 P_{e,a,max}\eta_{e,chrg}} - \frac{\sum_{j=t_{arrive}}^{t_{depart}} \varphi_{e,j}(d)}{P_{e,a,max}\eta_{e,chrg}} \right\rceil \tag{4.33}$$

In Equation (4.33), $\lceil . \rceil$ is the ceiling function and $SoC_{e,ini}$ is the initial SoC value of EV e, when it is plugged-in for charging. To calculate the number of instants an EV needs to charge from the grid, it subtracts the total charging instants required to achieve the desired SoC from the estimated number of PV charging instants (instants during which PV energy production can be consumed by an EV free of cost). Based on the information so far, a generalized EV optimal charging algorithm can be developed to manage its day-ahead charging under the PV uncertainty. Algorithm 4.12 presents this functionality.

In the presented Algorithm 4.12, $\hat{\varphi}_e$ is the vector of estimated instantaneous PV energy production, whereas $\tilde{\varphi}_e$ is the sampled instantaneous PV energy production vector from a normal distribution with mean $\tilde{\varphi}_e$. The estimation vector $\tilde{\varphi}_e$ is updated after each day based on the observed PV energy production vector $P_{e,PV}$. Each element of this vector $P_{e,PV}$ represents the instantaneous PV energy production $P'_{PV}(t)$. This information is used by an EV agent to calculate its $i_{e,req}(d)$ value, which in turn is necessary to decide the number of base arms in $S_e(d)$. Function $||.||_1$ represents the L1 norm and returns the total number of selected base arms here. Terms $I_{m,m}$, and 0_m stand for the $m * m$ identity matrix, and the null vector of length m, respectively.

Algorithm 4.12 CMAB-based decentralized day-ahead smart grid optimization (each EV)

Require: Total number of instants (base arms) m
Require: Learning rate $\beta \in \mathbb{R}^+$
$e \coloneqq$ EV agent index
$d \coloneqq$ Learning day index
$S_e(d) \coloneqq$ Selected super arm on day d
$i_{e,req}(d) \coloneqq$ Total number of required instants for grid charging on day d
$rew_e(S_e(d)) \coloneqq$ Reward vector for playing $S_e(d)$ super arm
$P_{e,PV} \coloneqq$ PV energy production vector
$\hat{\varphi}_e \coloneqq 0_m$; $Y \coloneqq I_{m,m}$; $z \coloneqq 0_m$
for $d = 1, 2, 3, \ldots$ **do**
    $\tilde{\varphi}_e \sim \mathcal{N}(\hat{\varphi}_e, \beta^2 Y^{-1})$
    Calculate $i_{e,req}(d)$ using Equation (4.33)
    Select $S_e(d)$ using a *learning strategy* s.t. $||S_e(d)||_1 = i_{e,req}(d)$
    Observe $rew_e(S_e(d))$ using the reward model in Equation (4.30)
    Observe $P_{e,PV}$ from Algorithm 4.11
    Update parameters of the selected *learning strategy*
    $Y \coloneqq Y + I_k I_k^\top$; $z \coloneqq z + P_{e,PV}$; $\hat{\varphi}_e \coloneqq Y^{-1} z$
end for

Note 4.4.2

It should be noted that the main objective of this thesis is not to design a novel PV forecaster, but to develop a decentralized system that would optimize the energy flows in real-time by tackling the inherent uncertainties present in any PV forecaster. Furthermore, one can also easily replace the utilized Bayesian learning-based PV forecaster with a more sophisticated deep learning-based forecaster in the proposed smart grid control algorithms, if desired. This change would not require any changes in other parts of the proposed decentralized smart grid control system.

To deal with *real-time uncertainties*, the system should operate in real-time as well. This can be done by enabling the learning agent to alter its base arms selection after every instant during the day. In this way, if any system variable is changed intra-day, the learning agent will be capable of adapting to these changes during the day as well. It should be noted that an EV learning agent can modify only those base arms in $S_e(d)$, which are yet to be played i.e., only the charging decisions of the future can be altered but not of the past. If $i_{e,req,t}(d)$ are the number of future instants required by EV e to charge from the grid at instant t, and $i_{e,chrgd,t}(d)$ are the total number of instants, till instant t, at which EV e has charged from the electrical grid on day d. Then for real-time operation, Equation 4.33 can be modified as follows:

Real-time PV uncertainty management model

$$i_{e,req,t}(d) = \left\lceil \frac{60E_{bat}(SoC_{e,depart} - SoC_{e,ini})}{||[m]_i||_1 P_{e,a,max}\eta_{e,chrg}} - \frac{\sum_{j=t_{arrive}}^{t_{depart}} \varphi_{e,j}(d)}{P_{e,a,max}\eta_{e,chrg}} \right\rceil - i_{e,chrgd,t}(d) \tag{4.34}$$

The key idea is that an EV learning agent will re-evaluate Equation (4.34) after each passing instant. Thus, if any of the system variable is updated, this will be reflected in Equation (4.34). Based on this modification, Algorithm 4.13 presents the pseudo-code of a multi-armed combinatorial multi-armed bandit algorithm for smart grids optimization in real-time.

Algorithm 4.13 CMAB-based decentralized real-time smart grid optimization (each EV)

Require: Total number of instants (base arms) m
Require: Learning rate $\beta \in \mathbb{R}^+$
$e :=$ EV agent index
$d :=$ Learning day index
$t :=$ Intra-day time index
$S_e(d) :=$ Selected super arm on day d
$i_{e,req,t}(d) :=$ Total required instants for grid charging after instant t on day d
$i_{e,chrgd,t}(d) :=$ Total instants already charged from the grid till instant t on day d
$rew_e(S_e(d)) :=$ Reward vector for playing $S_e(d)$ super arm
$P_{e,PV} :=$ PV energy production vector
$\hat{\varphi}_e := 0_m$; $Y := I_{m,m}$; $z := 0_m$
for $d = 1, 2, 3, \ldots$ **do**
    $\tilde{\varphi}_e \sim \mathcal{N}(\hat{\varphi}_e, \beta^2 Y^{-1})$
    $i_{e,chrgd}(d) := 0$
    for $t = 1, 2, 3, \ldots, m$ **do**
        Calculate $i_{e,req,t}(d)$ using Equation (4.34)
        Select $S_e(d)$ using a *learning strategy* s.t. $\sum_{j>t} S_{e,j}(d) = ||S_e(d)||_1 = i_{e,req,t}(d)$
        Observe t-th element of $rew_e(S_e(d))$ using reward model in Equation (4.30)
        Observe t-th element of $P_{e,PV}$ from Algorithm 4.11
        $i_{e,chrgd,t}(d) := i_{e,chrgd,t}(d) + \mathbb{1}[t \in S_e(d)]$
    end for
    Update parameters of the *learning strategy*
    $Y := Y + I_k I_k^\top$; $z := z + P_{e,PV}$; $\hat{\varphi}_e := Y^{-1} z$
end for

The last main type of uncertainty is the *opponents' actions uncertainty* i.e., uncertainty in the choice of super arms of other agents from one agent's perspective. This uncertainty is directly linked to the choice of learning strategy in Algorithm 4.12 or Algorithm 4.13. This *learning strategy* can be based on any methodology that handles

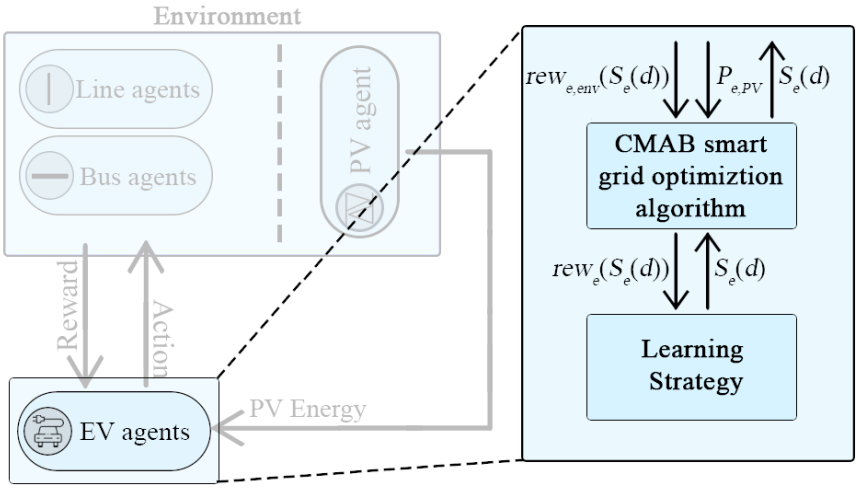

Figure 4.15: Interaction between the proposed CMAB-based MAS algorithm and the selected learning strategy, inside an EV agent.

the *exploration-exploitation* dilemma. Thus, the choice of learning strategy in the proposed system is crucial as a learning strategy with subpar performance under this type of uncertainty can lead to a higher regret. Interactions between the proposed Algorithm 4.13 and a learning strategy are shown in Figure 4.15. Most popular multi-armed bandit learning strategies choices are Thompson sampling (TS), upper confidence bound (UCB), and exponential-weight algorithm for exploration and exploitation (EXP3) etc.

Selfish Thompson sampling-based combinatorial multi-armed bandit algorithm has been shown to outperform both UCB-based and EXP3-based combinatorial multi-armed bandit strategies for IoT communication optimization [28]. However, as the problem at hand is optimal energy management in smart grids, thus all three popular learning strategies (i.e., Thompson sampling, UCB, and EXP3) are studied here. Each learning strategy is divided into two main sections. The first section is used to select the super arm to be played i.e., $S_e(d)$. Whereas, the second segment updates a set of parameters local to a learning strategy. Thompson sampling-based learning strategy is presented at first in Algorithm 4.14.

In the presented Thompson sampling-based learning strategy, κ is the learning parameter and $\hat{\theta}_e$ is the current estimate of the unknown parameters vector. Both *prior* and *posterior* are modeled as Gaussian distributions here. Term $\tilde{\theta}_e$ is the sampled vector of unknown parameters, which depends on $\hat{\theta}_e$. The learning parameter κ is used to control the Gaussian distribution's variance, and thus the degree of exploration. A higher value of variance in the estimate $\hat{\theta}_e$ corresponds to a higher degree of exploration. In the *action-selection* segment, the estimated optimal super arm on day d is picked for Algorithm 4.13. This super arm is played by the learning agent, and the reward vector is observed. Based on the observed reward vector, the learning strategy updates its estimate of $\hat{\theta}_e$ in the *update segment*. The learning strategy here can also be based on the UCB multi-armed bandit algorithm, Algorithm 4.5. The UCB-based learning strategy is given in Algorithm 4.15.

The presented UCB-based learning strategy comes from the CUCB algorithm [37]. Vector operations $[.]^{\circ -1}$ and $[.] \circ [.]$ stand for the Hadamard (element-wise) inverse and

Algorithm 4.14 Thompson sampling-based learning strategy

Require: Learning rate $\kappa \in \mathbb{R}^+$
Require: Total number of instants (base arms) from Algorithm 4.13 m
Require: Reward vector for the day from Algorithm 4.13 $rew_e(S_e(d))$
Require: Total required instants for grid charging from Algorithm 4.13 $i_{e,req,t}(d)$

$\hat{\theta}_e \coloneqq 0_m$
$A \coloneqq I_{m,m}$
$b \coloneqq 0_m$
▷ Action-selection segment
$\tilde{\theta}_e \sim \mathcal{N}(\hat{\theta}_e, \kappa^2 A^{-1})$
Select $S_e(d) = \arg\max_{S \in \{0,1\}^m} S^T \cdot \tilde{\theta}_e$ s.t. $\sum_{j>t} S_{e,j}(d) = ||S_e(d)||_1 = i_{e,req,t}(d)$
▷ Update segment
$A \coloneqq A + S_e(d) \cdot S_e(d)^\top$
$b \coloneqq b + rew_e(S_e(d))$
$\hat{\theta}_e \coloneqq A^{-1} b$

Algorithm 4.15 UCB-based learning strategy

Require: Learning day index from Algorithm 4.13 d
Require: Intra-day time index from Algorithm 4.13 t
Require: Total number of instants (base arms) from Algorithm 4.13 m
Require: Reward vector for the day from Algorithm 4.13 $rew_e(S_e(d))$
Require: Total required instants for grid charging from Algorithm 4.13 $i_{e,req,t}(d)$

$\hat{\theta}_e \coloneqq 0_m$
$N \coloneqq$ Vector containing number of times each base arm has been selected
$N \coloneqq 0_m$
▷ Action-selection segment
Select $S_e(d) = \arg\max_{S \in \{0,1\}^m} S^T \cdot \hat{\theta}_e + \sqrt{\frac{3}{2} \ln(d) N^{\circ -1}}$
s.t. $\sum_{j>t} S_{e,j}(d) = ||S_e(d)||_1 = i_{e,req,t}(d)$
▷ Update segment
$N \coloneqq N + S_e(d)$
$b \coloneqq b + rew_e(S_e(d))$
$\hat{\theta}_e \coloneqq N^{\circ -1} \circ b$

the Hadamard (element-wise) product respectively [72]. The philosophy behind *exploration* is the same as in the standard UCB algorithm. In the *action-selection* segment, term $\sqrt{\frac{3}{2}\ln(d)N^{\circ-1}}$ is associated with the degree of uncertainty in the estimated unknown parameters vector $\hat{\theta}_e(d)$. Thus, the super arm which maximizes the learning agent's expected reward, including this uncertainty term, is selected. In the *update segment*, $\hat{\theta}_e$ is updated based on the played super arm $S_e(d)$ and observed set of rewards $rew_e(S_e(d))$.

Next comes the EXP3-based learning strategy, which is inspired by the COMBAND algorithm [36]. Unlike the COMBAND algorithm, the presented EXP3-based learning strategy strives to maximize its reward instead of minimizing a loss function. Here, the idea is similar to the EXP3 algorithm, Algorithm 4.6. In the *action-selection* part, the learning strategy decides to perform exploration uniformly $\mathcal{U}$, with probability γ. Exploitation is done here, with $1-\gamma$ probability, by calculating the probabilities of each base arm and then selecting the super arm to be played $S_e(d)$ according to the calculated probability vector p. In the *update segment*, the weights of each base arm are updated based on the observed reward vector $rew_e(S_e(d))$. The studied EXP3-based learning strategy's execution steps are given in Algorithm 4.16.

Algorithm 4.16 EXP3-based learning strategy

Require: Total number of instants (base arms) from Algorithm 4.13 m
Require: Reward vector for the day from Algorithm 4.13 $rew_e(S_e(d))$
Require: Total required instants for grid charging from Algorithm 4.13 $i_{e,req,t}(d)$

1: $w := I_m$
▷ Action-selection segment
2: $p := (1-\gamma)w + \gamma\mathcal{U}$
3: Select a super arm $S_e(d)$ according to p s.t. $\sum_{j>t} S_{e,j}(d) = ||S_e(d)||_1 = i_{e,req,t}(d)$
▷ Update segment
4: $w := w \exp\left(\gamma(rew_e(S_e(d)) \left[\mathbb{E}\left[VV^T\right]\right]^+ S_e(d))\right)$, where V has law p, and $\left[\mathbb{E}\left[VV^T\right]\right]^+$ denotes the pseudo-inverse of $\left[\mathbb{E}\left[VV^T\right]\right]^+$

This concludes the modeling of the EV learning agent. In this section, all three main elements of the proposed reinforcement learning-based system are discussed in detail, i.e., the environment, the communication framework, and the learning agent. The designed system is completely decentralized because each network entity that runs into a problem (such as a transformer with congestion or a node with under-voltage) resolves it by sending messages to the flexible entities (in this case, the EVs), and each EV optimizes its charging strategy. The proposed system is real-time, model-free (i.e., it does not require an accurate distribution network model for its functionality), and scalable. The smart charging problem has been specifically studied in this chapter. However, the developed multi-agent system is fully capable of optimizing the operation of other electrical network elements as well such as household appliances. In the subsequent chapter, the transformation of this combinatorial multi-armed bandit learning-based multi-agent system to a combinatorial multi-armed bandit learning-based adaptive multi-agent system is presented to tackle the same optimization problem of smart charging under uncertainties. A detailed evaluation of the proposed sys-

tem is also done in the coming chapter.

Note 4.4.3

As each super arm holds a linear structure, the best super arm can be evaluated by the algorithm in $O(m)$ time. This calculation of the best super arm depends only on the total number of base arms m and not on the total number of agents in the decentralized system. This ensures that the computational time of each agent in a decentralized multi-agent system based on the proposed algorithm will remain the same (for a fixed m) irrespective of the number of agents in the system, and hence the system is scalable. Additionally, a larger electrical distribution network with a higher number of EV agents typically has a higher congestion limit as well, which is also expected to aid the scalability of the proposed decentralized smart grid optimization system.

4.5 Conclusion

The major drawback of the adaptive multi-agent system presented in Chapter 2, i.e., performance degradation due to lack of anticipative abilities, was addressed in this chapter. This was achieved through the amalgamation of the multi-agent system framework with combinatorial multi-armed bandit learning. At first, a detailed introduction to the multi-armed bandit class of reinforcement learning algorithms was provided. This was followed by an in-depth design of a real-time control system to optimize energy flows in a smart grid under uncertainties. The resulting system functioned in a decentralized manner and utilized the concepts of combinatorial multi-armed bandits to manage a number of stochasticities in the system. This system was designed to solve the smart charging optimization problem in real-time. The goal was to control the instantaneous charging power of each electric vehicle in real-time to minimize their daily charging cost while satisfying grid constraints, prosumer constraints, and maintaining fairness among electric vehicles. The studied problem involved two uncertainties, i.e., uncertainty in the daily PV energy production which can be utilized by electric vehicles free of cost, and uncertainty in the action selected by each opponent player from one electric vehicle agent's perspective. In the next chapter, this combinatorial multi-armed bandit multi-agent system will be transformed into a combinatorial multi-armed bandit adaptive multi-agent system, benefiting from the advantages of both combinatorial multi-armed bandit and adaptive multi-agent systems.

Chapter 5

Adaptive multi-agent multi-armed bandit system for smart charging under uncertainties

> Yesterday I was clever, so I wanted to change the world. Today I am wise, so I am changing myself.
>
> Rumi

Summary

This chapter integrates the decentralized control systems from Chapter 2 and Chapter 4, creating a hybrid system that combines the advantages of both approaches. The developed system utilizes adaptive multi-agent system theory to maintain scalability and adaptability, while also integrating combinatorial multi-armed bandit learning to enhance performance under uncertainties by incorporating anticipative decision-making capabilities. The chapter also includes a detailed evaluation of the developed system, comparing its performance with other baseline electric vehicle charging strategies through simulation-based experiments under stochastic conditions.

Contents

This chapter begins with a detailed explanation of the design of the proposed adaptive multi-agent multi-armed system, developed to optimize the previously introduced smart charging problem in Section 4.1. The distinctions between this system and the multi-agent multi-armed bandit system presented in the previous chapter are also detailed. Furthermore, three baseline electric vehicle charging strategies are discussed in Section 5.2. These baseline strategies serve as benchmarks to evaluate the performance of the system proposed in this chapter through simulation-based experiments in Section 5.3. These experiments incorporate stochasticity in the energy generated by solar photovoltaic panels, as described in Section 4.1. Finally, the chapter concludes with a summary in Section 5.4.

5.1 Proposed adaptive multi-agent multi-armed bandit system

In Chapter 2, it has been discussed how adaptability through self-organization of adaptive multi-agent systems can be more suitable to manage non-linear dynamic systems. Furthermore, it has also been discussed in Chapter 4 that combinatorial multi-armed bandit can help in tackling the uncertainties that may be encountered in real-life. This section presents a decentralized smart grid control system that combines the concepts of combinatorial multi-armed bandit with the framework of an adaptive multi-agent system. The resultant decentralized system is expected to optimally control complex electrical networks in the presence of real-life uncertainties. To formulate this system, the designs of already presented systems in Section 2.4 and Section 4.4 will be utilized here. Concisely, the proposed system here will follow the same adaptive multi-agent design described earlier in Section 2.4 but instead of utilizing rule-based AMAS for decision making, each EV will be using combinatorial multi-armed bandit for intelligent decision making.

System modeling

The mapping of physical electrical grid elements to software agents in the designed CMAB-based adaptive multi-agent system is the same as proposed earlier in Section 4.4. Each electrical line, electrical bus, electric vehicle and photovoltaic panel present in a distribution grid is mapped as an individual line agent, bus agent, electric vehicle agent, and photovoltaic agent. The objective of each agent type is also the same here i.e., a line agent is designed to protect against electrical current congestion, a bus agent is designed to prevent voltage limits violation, an electric vehicle agent is designed to minimize its daily charging cost while also helping other agent types in satisfying their constraints, and a photovoltaic agent is designed to communicate instantaneous PV energy production data to electric vehicles. All line, bus and EV agents hold a criticality value (between -1 and 1), and each of those agents tries to minimize its absolute criticality and the absolute criticality values of its neighboring agents at all instants. As the range of criticalities is between -1 and 1 here (and not between 0 and 1 as it was in Section 4.4), thus the *comparison of criticalities* principle is transformed

to the *comparison of absolute criticalities* principle. This principle is described as follows:

Comparison of absolute criticalities principle

Definition 5.1.1 (Comparison of absolute criticalities principle)**.** According to this principle, an agent compares the absolute value of its instantaneous criticality with the absolute instantaneous criticality values of its neighboring agents. Then, the instantaneous action made by this agent is to help the agent with the highest absolute instantaneous criticality.

Each agent goes through three stages (perception, decision, and action) during its single cycle while trying to achieve its own goal and helping its neighboring agents in achieving their goal(s). Line and bus agents in the designed AMAS belong to the collaborative class of agents, while each EV agent is an intelligent agent according to the Nwana's agent typology, shown in Figure 2.7. Detailed modeling of each agent type is presented next.

Line agent

The objective of each line agent is to keep the electric current, flowing through its corresponding physical electrical line, within its rated value. This can be achieved by an adaptive line agent by keeping its criticality value as close to zero as possible. The instantaneous line criticality value ranges between -1 (representing congestion due to a large power outflow towards the grid) and 1 (representing congestion due to a large power inflow from the grid). Thus, an instantaneous criticality value is calculated by each line agent using the following model:

Line agent's criticality model

$$Cr_{l,ab}(t) = \begin{cases} 0 & \text{if } |I_{ab}(t)| < I_{ab,th} \\ \frac{I_{ab}(t)}{|I_{ab}(t)|} \cdot \frac{|I_{ab}(t)| - I_{ab,th}}{I_{ab,max} - I_{ab,th}} & \text{if } I_{max} \geq |I_{ab}(t)| \geq I_{ab,th} \\ \frac{I_{ab}(t)}{|I_{ab}(t)|} & \text{if } |I_{ab}(t)| > I_{ab,max} \end{cases} \tag{5.1}$$

In Equation (2.18), $Cr_{l,ab}(t)$ is the instantaneous criticality of line agent corresponding to the electrical line connecting bus a and bus b, $I_{ab}(t)$ is the instantaneous electrical current flowing from bus a to bus b (positive in case of inflow from the grid and negative in case of outflow to the grid), $|I_{ab}(t)|$ is the absolute value of this instantaneous electrical current, $I_{ab,max}$ is the rated electrical current value through the electrical line between bus a and bus b, and $I_{ab,th}$ is a threshold value on the electrical current between bus a and bus b. The line criticality value $Cr_{l,ab}(t)$ is zero when the instantaneous line current is below the set threshold value, and it starts increasing linearly otherwise. The absolute value of this line criticality is maximum (i.e., 1) when its instantaneous electrical current is equal to or more than the rated value. This relationship between a line

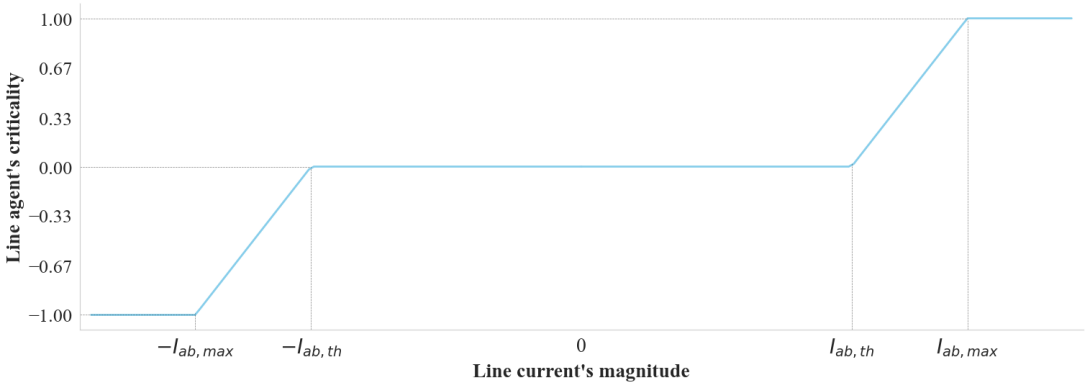

Figure 5.1: Relationship between line agent's criticality and electrical current's magnitude.

agent's criticality and its instantaneous electrical current is shown in Figure 5.1. The negative signs in Figure 5.1 with electrical current values indicate that the electrical current is flowing towards the grid. It can be verified that the line criticality is zero if the instantaneous electrical current remains within its threshold values. The line criticality starts increasing (or decreasing) linearly as soon as the instantaneous value of the electrical current goes above the set threshold value.

Similar to the system described in Section 2.4, a line agent also communicates only with its neighboring agents. This is one of the key differences between the system earlier proposed in Section 4.4 and this system. This dissimilarity is shown in Figure 5.2. The multi-agent system in the middle of Figure 5.2 follows the design rules of a "classical" multi-agent system proposed earlier in Section 4.4. It can be seen that each line agent will communicate with all bus agents of the system. However, this is not the case in the proposed adaptive multi-agent system in this section. As shown in Figure 5.2, the AMAS (on the right) only allows communication within the neighborhood

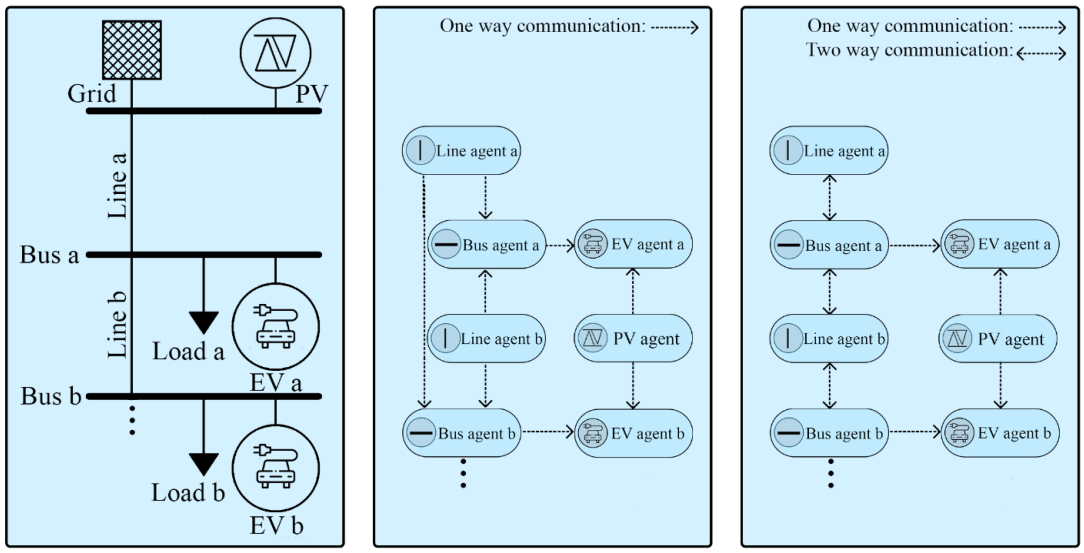

Figure 5.2: Section of a distribution network (left), its equivalent MAS (middle), and its equivalent AMAS (right).

of each line agent. The definition of the neighborhood of a line agent is the same as its definition in Section 2.4 (i.e., the neighborhood of a line agent that connects electrical bus *a* and electrical bus *b* consists of bus agents corresponding to electrical buses *a* and *b*). Furthermore, similar to the system in Section 2.4, a line agent here utilizes the *comparison of absolute criticalities* principle to decide if it should forward its criticality message or instead it should forward the received criticality message of one of its neighbors. For example, if line agent *b* in Figure 5.2 (right) is critical and it has also received a criticality message from bus agent *a* then it will compare its absolute criticality with the absolute criticality of the received request from bus agent *a* to decide which request it should forward to bus agent *b*. The flow of information here is also either upstream or downstream, i.e., the message received by line agent *b* from bus agent *a* in Figure 5.2 (right) can only be forwarded to bus agent *b* (and not back to bus agent *a*).

Unlike the system described in Section 2.4, a line agent does not send the same criticality message to all EVs requesting cooperation. Instead, similar to the CMAB-based MAS explained in Section 4.4, a line agent also uniformly samples a set of EV agents from the $[g] = 1, 2, 3, ..., g$ set of EVs which are charging at the same instant and causing current congestion. The set of EVs that have not been sampled is denoted by $[h] = 1, 2, 3, ..., h$. These EVs will not receive the line agent's cooperation request and thus they can charge simultaneously at their respective maximum power without causing electrical current congestion in the system. The uniformly sampled EV agents $[x] = [g] - [h]$ will receive the line agent's cooperation request. The criticality message sent by a line agent consists of a dual pair. The first element of this dual pair is the criticality value associated with the most critical agent determined through the *comparison of absolute criticalities* principle while the second element represents the set of EVs $[h]$ (uniformly sampled from all EVs present in the distribution network), picked by this most critical agent. The functionality of a line agent is described in Algorithm 5.1.

In Algorithm 5.1, a line agent goes through three stages during each cycle. These stages are perception, decision, and action. During the perception stage, a line agent calculates its instantaneous criticality value which depends on the observed value of its instantaneous current $I_{ab}(t)$. It also receives a set of messages from its neighboring agents. If this line agent is critical then it will uniformly sample a set of EVs $[x]$, which will be requested to perform cooperative actions. During the decision stage, each line agent is applying the *comparison of absolute criticalities* principle. If the criticality of this line agent is greater than its neighboring criticality then the set of EVs uniformly sampled by this line agent will be asked for cooperation. Otherwise, if a neighboring agent has a higher criticality, then the set of EVs sampled by that agent with a higher criticality is used for cooperation requests. Functions $\max_{Cr} \mathcal{R}(t)$ and $\arg\max_{Cr} \mathcal{R}(t)$ returns the maximum criticality in $\mathcal{R}(t)$ and the argument holding the maximum criticality in $\mathcal{R}(t)$ respectively. A line agent will decide to forward its criticality message to its neighboring agents if it has the highest absolute criticality value compared to all requests received in $\mathcal{R}(t)$. Otherwise, it will forward the message in $\mathcal{R}(t)$ with the highest criticality. Finally, during the action stage, the selected criticality message is forwarded to all neighboring agents.

Algorithm 5.1 CMAB-based AMAS line agent's functionality

Require: Electrical line's rated current $I_{ab,max}$
Require: Electrical line's threshold current $I_{ab,th}$
▷ **Perception stage**
$I_{ab}(t)$:= Perceived instantaneous line current from the sensor
$Cr(t)$:= Line agent's instantaneous criticality calculated using Equation (5.1)
$\mathcal{R}(t)$:= Set of requests received by line agent from its neighboring agents
$[x]$:= Uniformly sampled set of EVs for requesting cooperative actions
▷ **Decision stage**
Cr_f := Criticality value to be forwarded
$[x]_f$:= Information regarding $[x]$ to be forwarded
$Cr_f := 0$
$[x]_f$:= null
if $(|\max_{Cr} \mathcal{R}(t)| <= |Cr(t)|)$ **then**
    $Cr_f := Cr(t)$
    $[x]_f := [x]$
else
    $Cr_f := \max_{Cr} \mathcal{R}(t)$
    $[x]_f$:= Set $[x]$ corresponding to $\arg\max_{Cr} \mathcal{R}(t)$ request
end if
▷ **Action stage**
if $(Cr_f \neq 0)$ **then**
    Forward $(Cr_f, [x]_f)$ to neighboring agents
end if

Bus agent

The goal of each bus agent is to keep the magnitude of its instantaneous bus voltage within the specified limits. Similar to a line agent, this goal is achieved by an adaptive bus agent by keeping its criticality value as close to zero as possible. The instantaneous bus criticality value also ranges between -1 (in case of an over-voltage issue) and 1 (in case of an under-voltage issue). The instantaneous criticality value is calculated by each bus agent using the model presented in Equation (5.2). In Equation (5.2), $Cr_{b,a}(t)$ stands for the bus criticality of bus a and instant t. Terms $V^{-}_{a,th}$, $V^{+}_{a,th}$, $V_{a,min}$, and $V_{a,max}$ stands for the negative voltage threshold, the positive voltage threshold, the rated minimum voltage, and the rated maximum voltage at bus a, respectively. It should be noted that here $V_{a,min} < V^{-}_{a,th} < V^{+}_{a,th} < V_{a,max}$. The value of $Cr_{b,a}(t)$ is zero if the instantaneous voltage at bus a remains between $V^{-}_{a,th}$ and $V^{+}_{a,th}$. If this condition is violated then either an over-voltage issue or an under-voltage issue is present at the electrical bus. The bus criticality is 1 in case of an under-voltage issue, -1 in case of an over-voltage issue, and 0 in case of no issue.

Bus agent's criticality model

$$Cr_{b,a}(t) = \begin{cases} 0 & \text{if } V^{-}_{a,th} \leq V_a(t) \leq V^{+}_{a,th} \\ \frac{V_a(t) - V^{-}_{a,th}}{V_{a,min} - V^{-}_{a,th}} & \text{if } V_{a,min} \leq V_a(t) < V^{-}_{a,th} \\ -\frac{V_a(t) - V^{+}_{a,th}}{V_{a,max} - V^{+}_{a,th}} & \text{if } V_{a,max} \geq V_a(t) > V^{+}_{a,th} \\ 1 & \text{if } V_a(t) < V_{a,min} < V^{-}_{a,th} \\ -1 & \text{if } V_a(t) > V_{a,max} > V^{+}_{a,th} \end{cases} \tag{5.2}$$

The relationship between a bus agent's criticality and its instantaneous bus voltage magnitude is plotted in Figure 5.3. The bus agent's criticality value remains zero as long as the instantaneous bus voltage is within its lower threshold value $V^{-}_{a,th}$ and its upper threshold value $V^{+}_{a,th}$. As soon as the voltage magnitude starts going below the lower threshold $V^{-}_{a,th}$, the bus criticality starts increasing linearly and becomes equal

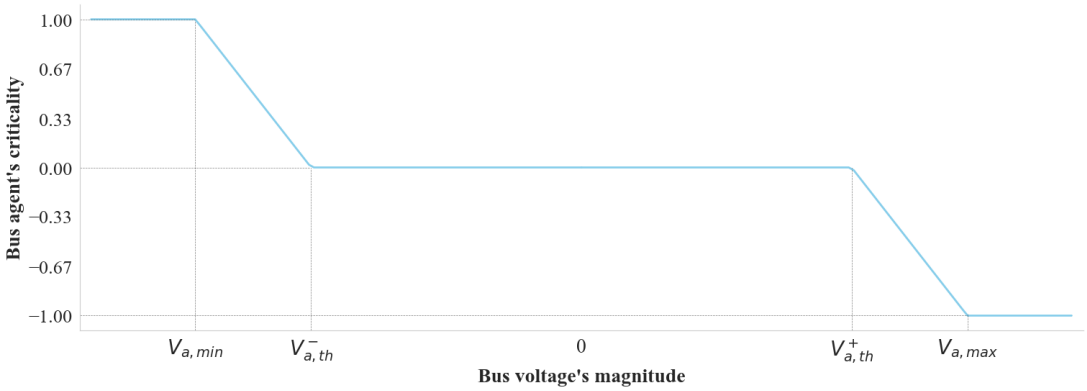

Figure 5.3: Relationship between bus criticality and instantaneous electrical bus voltage.

to 1 when the electrical bus starts facing an under-voltage issue. On the other hand, it starts decreasing linearly when the voltage magnitude starts going above the upper threshold value $V^{+}_{a,th}$. The bus criticality value reaches -1 when an over-voltage issue starts occurring at the electrical bus.

The functionality of a bus agent in this system is similar to that of a bus agent proposed earlier in Section 2.4. A bus agent here also communicates only in its neighborhood. For example, bus agent *a* in Figure 5.2 (right) can communicate only with line agent *a*, line agent *b*, and EV agent *a*. It must be noted that the bus agent does not always receive the criticality of the line. It only receives the line criticality when the value is non-zero. Furthermore, the information flow is also only in one direction here, i.e., bus agent *a* in Figure 5.2 (right) can forward the criticality message received from line agent *a* to only line agent *b* and EV agent *a*. However, there are also some similarities in the design of a bus agent here compared to that of a bus agent in Section 4.4. To give priority to the global line congestion in comparison to a local bus voltage limits violation, a bus agent will always forward the criticality message of a line agent to its neighboring EV agent irrespective of its criticality. In case no line agent criticality message has been received, a bus agent can forward its criticality message to its neighboring EV agent if it is critical. Algorithm 5.2 presents the functionality of a bus agent in this system.

Algorithm 5.2 CMAB-based AMAS bus agent's functionality

Require: Electrical bus' allowed minimum and maximum voltages $V_{a,min}$, $V_{a,max}$
Require: Electrical bus' threshold voltages $V^{-}_{a,th}$, $V^{+}_{a,th}$
▷ **Perception stage**
$V_a(t) :=$ Perceived instantaneous bus voltage from the sensor
$Cr(t) :=$ Bus agent's instantaneous criticality calculated using Equation (5.2)
$\mathcal{R}(t) :=$ Set of requests received by line agent from its neighboring agents
▷ **Decision stage**
$Cr_f :=$ Criticality value to be forwarded
$[x]_f :=$ Information regarding $[x]$ to be forwarded
$Cr_f := 0$
$[x]_f :=$ null
if (cooperation request due to line congestion is received) **then**
    $Cr_f := \max_{Cr} \mathcal{R}(t)$
    $[x]_f :=$ Set $[x]$ corresponding to $\arg\max_{Cr} \mathcal{R}(t)$ request
else
    if $(Cr(t) \neq 0)$ **then**
        $Cr_f := Cr(t)$
    end if
end if
▷ **Action stage**
if $(Cr_f \neq 0)$ **then**
    Forward $(Cr_f, [x]_f)$ to neighboring agent(s)
end if

During the perception stage in Algorithm 5.2, a bus agent is observing the instanta-

neous value of its bus voltage and calculating its instantaneous criticality based on the observed value. It is also receiving criticality message(s) from neighboring agents. It is evident that during the decision stage priority is given to solving the line congestion issue. This is because line congestion is expected to have a more global impact on a distribution network compared to a local bus voltage limits violation. However, one can easily design a system in which equal priority is given to both issues. A bus agent in such a system would apply the *comparison of absolute criticalities* principle to decide which criticality message should be forwarded instead of giving priority to solving a line congestion issue. Finally, a bus agent communicates with its neighboring agents in the action stage.

PV agent

The functionality of a PV agent is the same as it has been described in Algorithm 4.11, i.e., a PV agent is communicating the magnitude of instantaneous energy generation by its PV panel to all EV agents in the system. This communicated information is used by each EV agent to learn the trend of daily PV energy generation through Bayesian learning.

EV agent

Each EV agent is a learning agent here as well similar to the system in Section 4.4. It is interacting with its environment. However, an EV agent's interactions in this system are slightly different compared to their interactions in the MAS of Section 4.4. This is because line and bus agents are communicating their criticality values (and not reward values) in this adaptive multi-agent system. Thus, the RL learning model in Figure 4.12 can be transformed into the RL learning model presented in Figure 5.4 for this learning-based adaptive multi-agent system.

Due to this difference in agent's interactions, the reward model given in Equation (4.30) needs to be transformed into a reward model involving criticalities. This must be done as an EV agent is a combinatorial multi-armed bandit-based learning agent in the proposed system as well, and generally in such a learning system the concept of reward is used to evaluate the optimality of an agent's instantaneous action. Thus, the criticality received from the environment should be transformed into a reward value here. The following EV agent's criticality-based reward model is used for that purpose:

EV learning agent's crticality-based reward model

$$rew_{e,i}(S_e(d)) = \begin{cases} -Cr_{e,env,i}(S_e(d)) & \text{if } Cr_{e,env,i}(S_e(d)) \neq 0 \\ 1 - c(i) & \text{if } Cr_{e,env,i}(S_e(d)) = 0 \end{cases} \tag{5.3}$$

In Equation (5.3), term $Cr_{e,env,i}(S_e(d))$ stands for the criticality observed by EV agent e, on day d, at instant $i \in S_e(d)$, from its environment (i.e., $Cr_{e,env,i}(S_e(d))$ is the criticality value received from Algorithm 5.2). Whereas, $rew_{e,i}(S_e(d))$ represents the reward of EV agent e, on day d, at instant $i \in S_e(d)$. This reward is calculated based

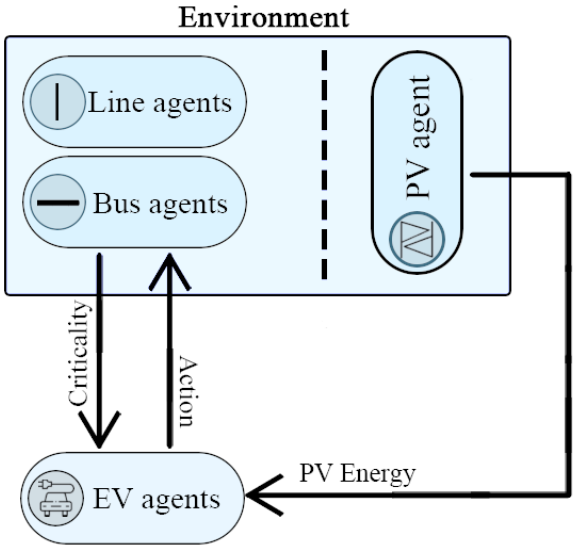

Figure 5.4: Reinforcement learning model of the proposed AMAS.

on the received criticality information and the instantaneous electricity price $c(i)$. If there is congestion in the environment (i.e., $Cr_{e,env,i}(S_e(d)) \neq 0$) then the observed reward depends on the criticality value obtained from the environment. It must be remembered that $S_e(d)$ stands for the super arm picked by EV agent e, on day d.

The rest of the internal working of an EV agent is the same as it was in Section 4.4, i.e., an EV agent utilizes combinatorial multi-armed bandit algorithm to minimize its daily charging cost while also satisfying a set of grid and prosumers constraints. Each day d is divided into $[m] = 1, 2, 3, ..., m$ equally spaced instants (also known as base arms). Each EV agent either selects an instant to charge from the grid or does nothing at that instant, i.e., the action set of an EV agent consists of two values here as well $0, 1$. Each EV selects a set of base arms (i.e., a super arm) to charge during the remainder of its connection time with the grid to achieve the desired SoC before its departure time. This selected super arm can be updated in real-time to handle the real-time uncertainties discussed earlier in Section 4.4 when the design of a CMAB-based learning EV agent was presented in detail. The functionality of a CMAB-based AMAS EV agent is given in Algorithm 5.3.

A CMAB-based EV agent also goes through three AMAS stages in Algorithm 5.3, i.e., perception, decision, and action. First, an EV agent is selecting a super arm to play using combinatorial multi-armed bandit learning during the decision stage. Information regarding the number of instants already charged from the grid $i_{e,chrgd,t}(d)$ on day d is used by an EV agent to calculate the number of instants required to charge from the grid $i_{e,req}(d)$ using Equation (4.34). This $i_{e,req}(d)$ value is then used to calculate the optimal super arm to be played. The methodology to calculate this optimal super arm depends on the learning strategy utilized (Thompson Sampling, UCB, or EXP3 as discussed in the previous chapter). Each EV agent sets the instantaneous charging power of its corresponding EV based on the selected super arm in the action stage. Finally, an EV agent will observe instantaneous criticality value from the environment $Cr_e(S_e(d))$ and it will calculate its instantaneous reward $rew_e(S_e(d))$ based on the ob-

Algorithm 5.3 CMAB-based AMAS EV agent's functionality

Require: Total number of instants (base arms) m
Require: Learning parameter $\beta \in \mathbb{R}^+$
$e \coloneqq$ EV agent index
$d \coloneqq$ Learning day index
$S_e(d) \coloneqq$ Selected super arm on day d
$i_{e,req}(d) \coloneqq$ Total number of required instants for grid charging on day d
$i_{e,chrgd,t}(d) \coloneqq$ Total instants already charged from the grid till instant t on day d
$rew_e(S_e(d)) \coloneqq$ Reward vector for playing $S_e(d)$ super arm
$Cr_e(S_e(d)) \coloneqq$ Instantaneous environment criticality
$P_{e,PV} \coloneqq$ PV energy production vector
$\hat{\varphi}_e \coloneqq 0_m$; $Y \coloneqq I_{m,m}$; $z \coloneqq 0_m$
for $d = 1, 2, 3, ...$ **do**
    $\tilde{\varphi}_e \sim \mathcal{N}(\hat{\varphi}_e, \beta^2 Y^{-1})$
    $i_{e,chrgd}(d) \coloneqq 0$
    for $t = 1, 2, 3, ..., m$ **do**
▷ **Decision stage**
        Calculate $i_{e,req,t}(d)$ using Equation (4.34)
        Select $S_e(d)$ using a *learning strategy* s.t. $\sum_{j>t} S_{e,j}(d) = ||S_e(d)||_1 = i_{e,req,t}(d)$
▷ **Action stage**
        Set EV's instantaneous charging power
▷ **Perception stage**
        Perceive $Cr_e(S_e(d))$ from the environment
        Observe t-th element of $rew_e(S_e(d))$ using reward model in Equation (5.3)
        Observe t-th element of $P_{e,PV}$ from Algorithm 4.11
        $i_{e,chrgd,t}(d) \coloneqq i_{e,chrgd,t}(d) + \mathbb{1}[t \in S_e(d)]$
    end for
    Update parameters of the *learning strategy*
    $Y \coloneqq Y + I_k I_k^\top$; $z \coloneqq z + P_{e,PV}$; $\hat{\varphi}_e \coloneqq Y^{-1} z$
end for

served environment criticality. Each EV agent is also observing the instantaneous PV energy production data from PV agent(s). Bayesian learning is used by EV agent e to make its estimation of the instantaneous PV production $\hat{\varphi}_e$ based on the observed data. Also, learning parameters corresponding to PV energy generation estimation (Y, z, and $\hat{\varphi}_e$) and the selected learning strategy are updated based on the observed $P_{e,PV}$, at the end of each day d.

Note 5.1.1

Two crucial points regarding the proposed system should be emphasized:

- The proposed system is designed to be theoretically scalable, as each electric vehicle agent can evaluate its estimation of the best super arm in $O(m)$ time, where m represents the total number of available base arms. This property is ensured because the super arms played by each electric vehicle agent are assumed to follow a linear structure, denoted as $\mathbb{E}\left[r_\mu(S_e(d))\right] = (S_e(d))^T \cdot (\theta_e(d))$.
- The system design ensures that critical private data of each agent type is not required to be shared with other agents in the system. Only instantaneous criticality values need to be shared. Additionally, no direct communication between electric vehicle agents is necessary for the proposed system, eliminating the need to share sensitive information related to electric vehicles (e.g., arrival time, departure time, desired state of charge, etc.) with other vehicles in the system.

This completes the modeling of the presented system which combines combinatorial multi-armed bandit learning with the framework of adaptive multi-agent systems. The proposed system is fully decentralized, real-time, model-free, scalable as well as adaptable. To evaluate the performance of this CMAB-based adaptive multi-agent system simulation studies are performed with the assumptions that communications among agents are synchronous as well as the speed of communication is much higher (in milliseconds) compared to the temporal resolution of decision-making in the system (one minute). In the following sections, the simulation-based experiments conducted to evaluate the performance of the proposed combinatorial multi-armed bandit-based adaptive multi-agent system (CMAB-based AMAS) are presented. The objective is to statistically demonstrate that the CMAB-based AMAS can effectively control large-scale smart grids in real-time and yield near-optimal solutions even in the presence of real-life uncertainties.

5.2 Baseline EV charging strategies

To evaluate the performance of the proposed adaptive multi-agent combinatorial multi-armed smart charging system, a number of baseline charging strategies are highlighted in this section. The solutions obtained through these charging strategies will be used

in quantifying the improvements made through multi-armed bandit learning in Section 5.3. The following charging strategies are selected as baselines:

- Uncontrolled charging strategy
- Centralized MILP charging strategy
- CMAB-based adaptive multi-agent charging strategy (no PV estimation)

Uncontrolled charging strategy:

Uncontrolled charging is one of the most commonly used practical EV charging strategies [95]. This charging strategy has already been described in Algorithm 3.1. In an uncontrolled charging strategy, EV e starts charging at its rated power $P_{e,a,max}$ as soon as it is plugged-in by the EV owner. Evidently, it is not an optimal charging strategy. This is because EVs are charging as soon as they are connected to the electrical grid. These EVs do not take into account the variability in daily electricity pricing. Furthermore, the impact of each EV's charging on the grid is not observed by an EV either. Thus, all EVs will continue to be non-optimal and may compromise the electrical grid's stability due to peak load demands. This charging strategy does not handle the PV energy production uncertainty (and thus does not benefit from freely available PV energy), the real-time uncertainty, and the opponents' actions uncertainty.

Centralized MILP charging strategy:

The centralized MILP optimization approach already discussed in Chapter 3 (in Section 3.1) can be applied here as well to solve the smart charging optimization problem. However, in the standard MILP formulation, the decision variable x belongs to the set of non-negative integers Z^+ (Equation (3.2)). Furthermore, in MILP it is assumed that the objective function, Equation (3.1), and the problem constraints, Equation (3.2), are linear. In the studied smart charging problem non-negative instantaneous EV charging power is assumed i.e., $P_{e,a}(t) \in [0, P_{e,a,max}]$. However, the originally presented smart charging problem belongs to the quadratic constrained programming (QCP) class of optimization problems [4]. The objective function in Equation (4.1) involves an absolute function term, which is not a linear function. Furthermore, there exists a product of voltages in Equation (4.5), which again is not a linear function. Thus, to apply MILP optimization, linearization of the mentioned equations must be performed.

Objective function linearization

Linearization of the original non-linear smart charging objective function, Equation (4.1), is performed here. The non-linear absolute term is linearized by assuming the absolute term equal to a normal variable Q, and then putting constraints on this assumed variable. This is shown as follows:

Smart charging problem's linearized objective function

Objective function:

$$\min \sum_{e=1}^{E} C_e(d) = \min \sum_{e=1}^{E} \sum_{t=1}^{n} c(t) P_{e,a}(t) \Delta t - \sum_{e'=1}^{E} \sum_{e=e'}^{E} |C_{e',pu}(d) - C_{e,pu}(d)| \tag{5.4}$$

is equivalent to:

$$\min \sum_{e=1}^{E} C_e(d) = \min \sum_{e=1}^{E} \sum_{t=1}^{n} c(t) P_{e,a}(t) \Delta t - \sum_{e'=1}^{E} \sum_{e=e'}^{E} Q \tag{5.5}$$

subject to the constraints:

$$(C_{e',pu}(d) - C_{e,pu}(d)) \leq Q \tag{5.6}$$

$$(C_{e,pu}(d) - C_{e',pu}(d)) \leq Q \tag{5.7}$$

Constraint linearization

Now that the objective function has been linearized, the non-linear constraint in Equation (4.5) can also be linearized. In fact, the linearization of this constraint has already been described in Section 3.1 when the centralized MILP baseline has been discussed. Thus, the original non-linear constraint in Equation (4.5) can be replaced with the following two separate linear power flow equations:

Linearized power flow costraints

$$P_{ab}(t) = G_{ab}(t) \left(V_a(t) - V_b(t)\right) + B_{ab}(t) \left(\psi_a(t) - \psi_b(t)\right) \tag{5.8}$$

$$Q_{ab}(t) = B_{ab}(t) \left(V_a(t) - V_b(t)\right) + G_{ab}(t) \left(\psi_b(t) - \psi_a(t)\right) \tag{5.9}$$

A feasible set, given below, is formed to solve the smart charging problem as a mixed integer linear programming (MILP) optimization problem.

The uncertainty in opponents' actions does not apply here, as it is a centralized optimization strategy. Real-time uncertainty can be tackled by centralized MILP optimization, if it would be able to perform optimization in real-time. However, it will be shown in Section 5.3 that for a large-scale system, real-time solutions may not be possible. Finally, the presented MILP formulation requires PV forecasts to perform optimization. Thus, the uncertainty in PV energy production is not managed by this smart charging strategy. An error in PV forecast input co-relates directly to the ob-

Feasible set 5.1: linearized smart charging formulation

Objective function in Equation (5.5)

DSOs constraints in Equations (4.8)-(4.11)

Prosumers constraints in Equations (4.12)-(4.14)

Objective function linearization constraints in Equations (5.6)-(5.7)

Network's physical constraints in Equations (4.3)-(4.7) & (5.8)-(5.9)

tained solution's quality.

CMAB-based charging strategy (no PV estimation)

This charging strategy is a variation of the proposed adaptive multi-agent combinatorial multi-armed bandit smart charging presented in Section 5.1. In Algorithm 5.3, if the estimated PV energy production vector $\hat{\varphi}_e$ is assumed to be null at all instants, then it will give the CMAB-based adaptive multi-agent smart charging strategy (no PV estimation). Similar to Algorithm 5.3, this charging strategy satisfies all constraints of the smart charging problem under study. However, unlike Algorithm 5.3, this charging strategy does not try to learn any information related to daily PV energy production. As a baseline strategy, this EV charging strategy is included to see the impact of learning the daily PV energy production trend on the total optimization cost (sum of daily costs of all EVs). This will allow us to quantify the improvement (in terms of total daily charging cost) made by Algorithm 5.3 through learning the mentioned PV production trend. This charging strategy manages the real-time uncertainty as well as the uncertainty in opponents' actions, but it does not tackle the uncertainty in free-of-cost PV energy production.

5.3 Stochastic simulation-based experimentation

Simulation studies are performed to evaluate the performance of our proposed adaptive multi-agent combinatorial multi-armed bandit-based smart charging system. The details of each simulation study and its results are presented in this section. The presented results also include solutions obtained through all baseline charging strategies, mentioned in Section 5.2.

Simulation-based experimentation settings

Each performed simulation study includes modeling of an electrical distribution grid, careful selection of system variables ($P_{e,a,max}$, $E_{e,bat}$, $SoC_{e,max}$, etc.), data engineering, and software implementation of the earlier discussed charging strategies. Here, all these aspects are presented in detail.

Electrical distribution networks

Two simulation case studies are performed. The classification is made based on the size of the electrical distribution network used in these studies:

- Small-scale case study
- Large-scale case study

The *small-scale case study* consists of 55 load buses. Each of these load buses has a household and an EV attached to it. Thus, in total, the small-scale distribution network has 55 households and 55 EVs. The single-line electrical diagram of the small-scale network is shown in Figure 5.5. This distribution network comprises only a single district, which is connected to the grid side through an 11/0.4 kV grid transformer. This district consists of only one sub-district. This only sub-district is modeled as the IEEE low voltage test feeder (IEEE LVTF) [159]. The grid side also includes PV power stations in the studied system. A relatively smaller-scale case study, with 55 EV agents, is modeled here to conduct an in-depth optimality analysis of our proposed adaptive multi-agent CMAB-based smart charging system. Centralized charging strategies can provide an optimal solution, which can be used as the lower bound to study other charging strategies. However, centralized charging strategies may not scale well. As a result, the optimal solution will be unknown, and the optimality of different charging strategies cannot be studied. Thus, the main reason to perform this small-scale case study is its capability to provide the optimal solution when the centralized MILP optimization smart charging strategy is applied to it.

The aim of this thesis, however, is to actually design an optimal decentralized smart charging system for *large-scale* smart grids. Thus, a second case study is performed on a large-scale distribution network. This is termed as the *large-scale* case study

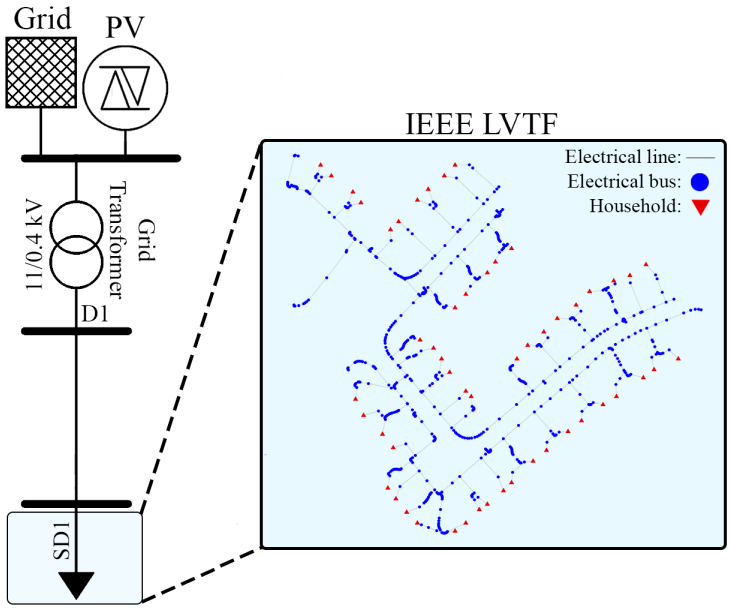

Figure 5.5: Distribution network used to perform the small-scale case study.

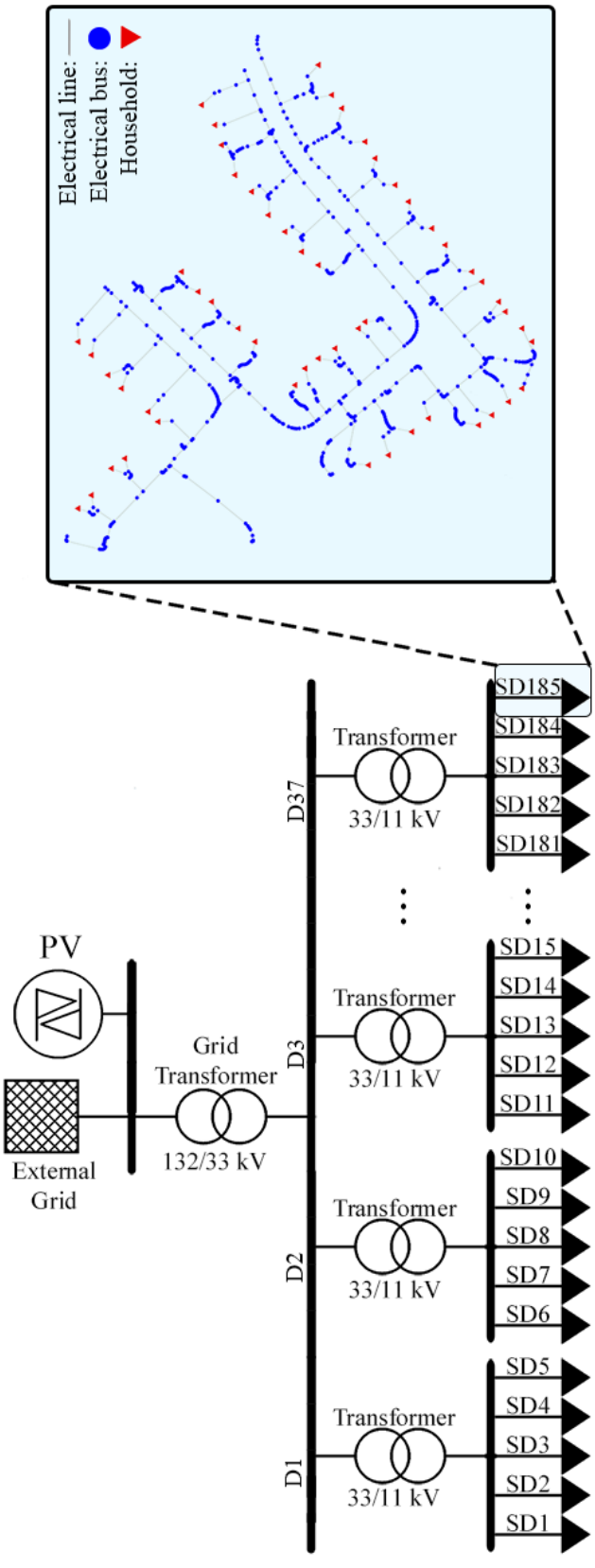

Figure 5.6: Distribution network used to perform the large-scale case study.

here. The single-line diagram of the studied large-scale distribution network is shown in Figure 5.6. The large-scale distribution network consists of 37 districts, connected to the grid side through 132/33 kV grid transformers. Each of these districts is further divided into five sub-districts. Thus, there are 185 total sub-districts in the modeled large-scale distribution network. These sub-districts are connected to their specific district through 33/11 kV transformers. Similar to the small-scale case study, each sub-district is modeled as the IEEE low voltage test feeder (IEEE LVTF) [159]. Each IEEE LVTF sub-district includes 55 load buses. Thus, there are a total of 10,175 households and 10,175 EVs in the studied electrical distribution network. The voltage magnitude at the grid bus is fixed to 1 per-unit. However, the grid generator's active and reactive powers are not set to specific values. These active and reactive powers would depend on the demanded (or generated) power by the connected distribution network.

Parameters

Each day is divided into 1440 equally spaced decision-making instants, i.e., each EV has to determine its instantaneous charging power at each minute of the day. Each EV's battery's minimum and maximum SoC values are set to 0.3 and 0.8, respectively. Between 0.3 and 0.8 is the ideal SoC range to decelerate electrolyte degradation and capacity loss in EV batteries [55]. The rated charging power of each EV $P_{e,a,max}$ is set to 7 kW. The charging/discharging efficiency $\eta_{e,a}$ of 0.95 is considered for each EV. To incorporate the heterogeneity introduced by different EV models in practical life, the battery capacity of each EV is modeled to hold the value belonging to any one of the following (Table 5.1) top-selling electric vehicles [98]:

EV model	Battery capacity (kWh)
Tesla Model 3	57.5
Renault Zoe	52
Peugeot 208 EV	51

Table 5.1: Battery capacities of top-selling EV models in France (January - December 2021).

The maximum $V_{a,max}$ and minimum $V_{a,min}$ voltage magnitudes allowed at each bus are 1.05 and 0.95, respectively. The rated current of each electrical line $I_{ab,max}$ depends on the model of the electrical line in the IEEE LVTF network, which comes along with the IEEE LVTF network model [159].

Datasets

Three datasets are used to model three main time-series elements of the system:

- Household loads data
- Photovoltaic irradiance data
- Electric vehicles data

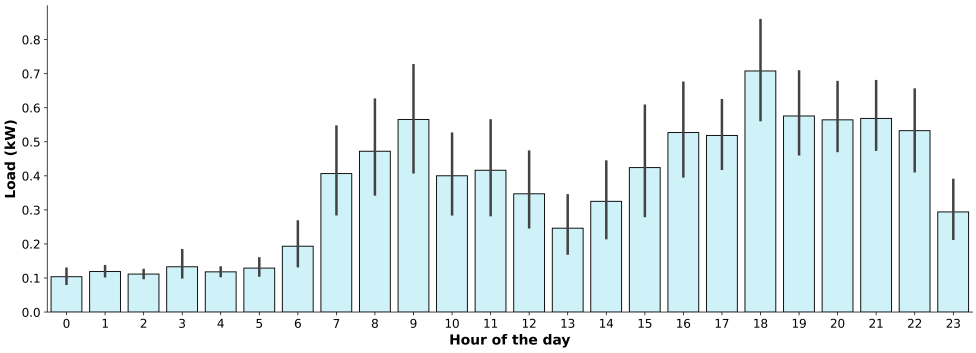

Figure 5.7: Sum of all load profiles in the IEEE LVTF dataset.

First, the *load profile* of each household load is provided along with the IEEE LVTF network model. The resolution of each household's load profile is 1 minute. The bar chart of the mean of all load profiles is presented in Figure 5.7. It can be seen that the accumulated load profile has a peak during the later hours of the day. Whereas, during the early hours of the day, this accumulated load is around its minimum value. This accumulated load directly corresponds to the distribution network's stress. The stress is maximum in the evening and minimum during the early morning.

Second, the PV irradiance data is obtained by the National Renewable Energy Laboratory (NREL) dataset [133]. The irradiance data utilized here corresponds to the location with 39.78 latitude and -105.23 longitude values in the NREL database for the year 2020. The resolution of this data is 1 minute as well. The heatmap of this irradiance dataset is shown in Figure 5.8. This heatmap plots the irradiance values during each day of the year against the irradiance values during each hour of the day. It can be seen in the heatmap that there are uncertainties in this instantaneous PV irradiance value. The direct impact of these uncertainties on the studied objective function in Equation (4.1) encourages each EV to make its decisions based on its estimation of the instantaneous PV production. The instantaneous PV energy production P_{PV} is linked to the instantaneous solar irradiance value $Irr(t)$ as follows:

$$P_{PV} = Irr(t) A \eta_{PV} \tag{5.10}$$

where A and η_{PV} represent the total area and the efficiency of PV panels, respectively.

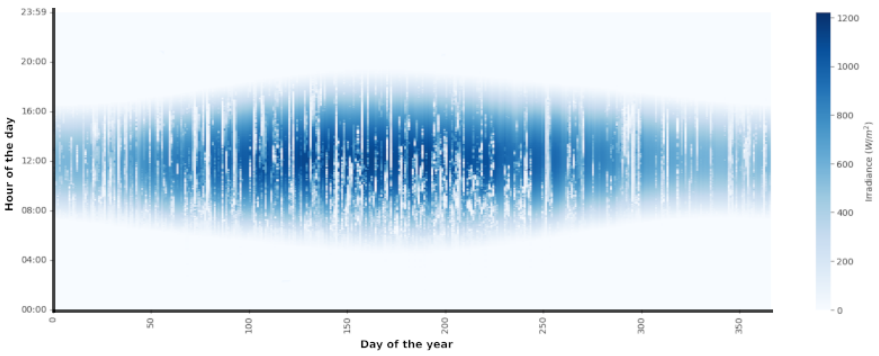

Figure 5.8: Heatmap of the NREL PV irradiance dataset.

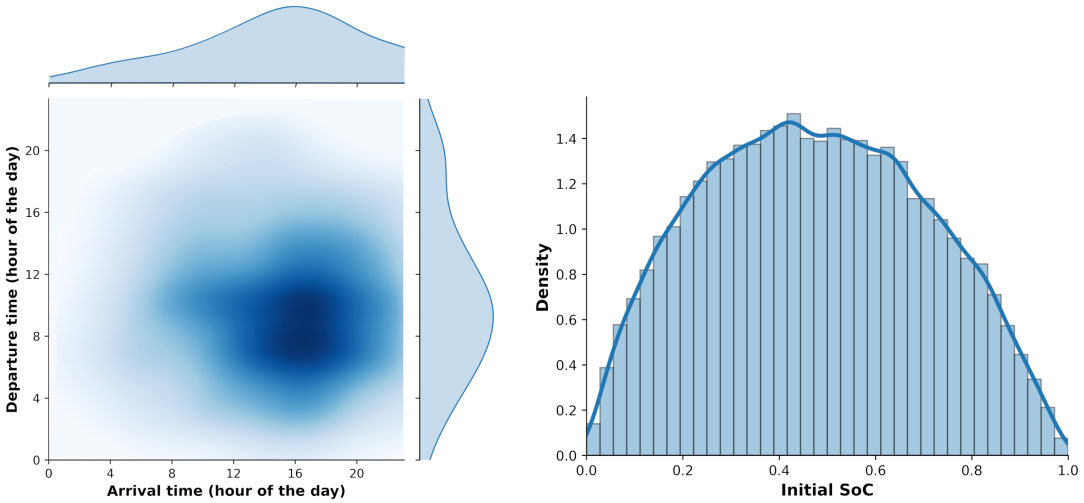

Figure 5.9: Jointplot of the arrival time distribution, departure time distribution (left), and the distribution of the initial SoC.

Third, the *electric vehicles'* dataset is required to model the arrival time, the departure time, and the initial state of charge of each electric vehicle. This dataset comes from the "TestanEV" project [41]. In this project, the mobility data of 185 electric vehicles were recorded over a period of one year. This data is utilized to model EVs in our experiments. The large-scale case study involves 10,175 EVs, whereas there are only 185 EVs in the dataset. Hence, a distribution is fitted to each required data variable (arrival time, departure time, and initial SoC) in the dataset. These distributions are shown in Figure 5.9. It can be seen that the arrival time's distribution has its peak around evening time. Around this time, the household load stress on the grid is at maximum as well, as shown in Figure 5.7. Thus, a demand peak can be formed, which can cause congestion in the network. It can also be seen that the departure time's distribution peaks in the morning hours, when people may leave for work, school etc. The EV load demand can be shifted to early hours to avoid the earlier discussed peak load demand. The initial state of charge distribution is also shown in Figure 5.9. Our simulation-based experimentation uses these distributions to sample required arrival times, departure times, and states of charge.

Implementation

All of the studied charging strategies are implemented in Python [181]. PandaPower is a Python library that can be used to model any desired distribution network and execute load flows [172]. This library is used by the uncontrolled EV charging Algorithm 3.1 to get the required power flow results. The centralized MILP optimization charging strategy does not require PandaPower, as the load flow equations have been modeled as hard constraints i.e., Equation (4.3)-(4.7). The CVXPY python-embedded modeling language is also utilized here to solve the centralized MILP optimization problem [51].

The working of combinatorial multi-armed bandit-based systems is shown in Figure 5.10. This system has two main components, i.e., the CMAB-based AMAS and the simulator. This CMAB-based smart charging system consists of an environment

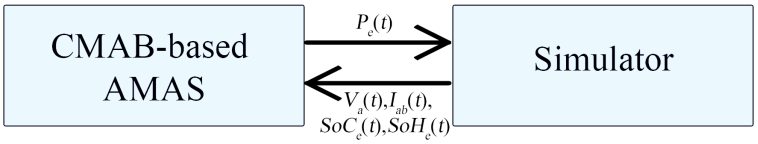

Figure 5.10: Block diagram of CMAB-based EV charging strategies implementation in Python.

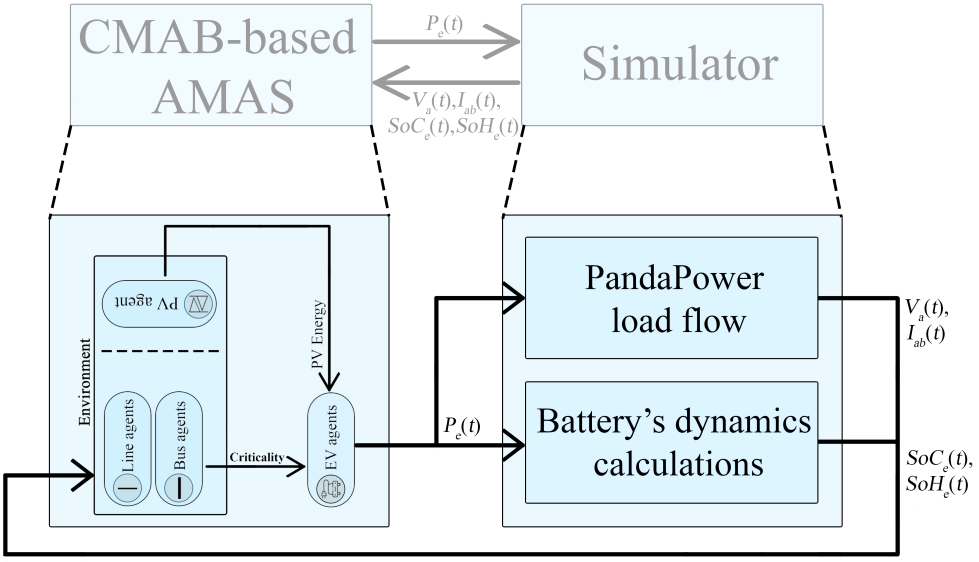

Figure 5.11: Expanded block diagram of CMAB-based EV charging strategies implementation in Python.

and EV agents. The simulator in Figure 5.10 is the new element here. This simulator is designed specifically to carry out simulation studies. The role of this simulator is basically to model the studied distribution network, perform power flows, and update the states of EVs (such as SoC and SoH). The expanded block diagram of system implementation is given in Figure 5.11.

Computing machine's specifications

All of the presented experiments are performed on a computing machine with the following specifications (Table 5.2):

	Specification
Processor	AMD Ryzen Threadripper 3970X (4.5 GHz)
Memory	DDR4 (128 GB, 3200 MHz)
Storage	SSD (1 TB, 560/530 MB/s)

Table 5.2: Specifications of the computing machine used to perform simulation-based experiments.

Evaluation metrics

In both of the described experiments, four key metrics are used to draw comparisons among different studied charging strategies. These metrics include *optimality*, *constraints satisfaction*, *fairness*, and *scalability*. The aim would be to compare the proposed CMAB-based adaptive multi-agent smart charging strategy with each of the discussed baseline charging strategies, in Section 5.2, using these metrics. This comparison would allow a better understanding of each EV charging strategy in realistic simulation scenarios.

Constraints satisfaction

The studied set of constraints i.e., Equation (4.3)-(4.14), must be satisfied as well. Thus, different EV charging strategies can be compared depending on whether these constraints are satisfied. Due to peak demand, one can expect the congestion constraints to be violated in uncontrolled EV charging. However, all optimization-based smart charging algorithms must satisfy this set of constraints.

Optimality

Another important aspect of an optimization problem's solution is its optimality. It is desired that the solution obtained through any smart charging strategy minimizes (or maximizes) the studied objective function. The objective here is to minimize electric vehicles' daily charging costs. The proposed CMAB-based adaptive multi-agent smart charging strategy (or any other EV charging strategy) will provide a solution to minimize this cost. However, whether this provided solution is optimal remains a question if no comparisons are made. Hence, the centralized MILP smart charging strategy is used to obtain optimal solutions. Then the evaluation of other EV charging strategies is made based on the obtained centralized MILP solutions.

This centralized optimal solution is termed as the *lower bound* in this simulation-based experimentation section. The studied EV charging strategies can match this lower bound at best. The centralized MILP solution depends on the system's temporal resolution, i.e., the length of each decision instant in a day (hour, minute, second, etc.). A relatively lower temporal resolution (e.g., hour) would mean fewer decision-making instants, directly corresponding to a lower-quality solution. Furthermore, centralized MILP optimization also requires PV energy production forecasts. This PV energy production forecast error correlates inversely to the obtained solution's quality. The heatmap plotting the impact of these two variables (temporal resolution and PV energy production forecast error) on the centralized MILP optimization performance is shown in Figure 5.12. The small-scale distribution network with earlier described datasets is used to obtain this plot. The difference (%) is calculated in Figure 5.12 by taking the point with null forecast error and 1-minute temporal resolution as the initial value.

It can be seen in Figure 5.12 that the optimization cost decreases as the temporal resolution decreases. This decline in the optimization cost is sharp when the temporal resolution is changed from hours to minutes. However, this reduction in the optimization cost plateaus when the temporal resolution is below five minutes. Thus, the centralized MILP optimization's temporal resolution will be fixed to 1-minute in the

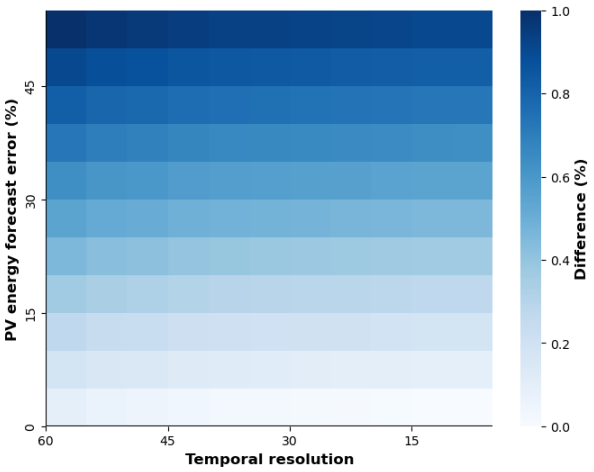

Figure 5.12: Impact of MILP's temporal resolution and forecast error on the obtained solution.

experiments. Also, it is evident in Figure 5.12 that the optimal solution can only be obtained when the PV energy production forecast is error-free. Hence, in the next section, the centralized MILP optimization solution with a 1-minute temporal resolution and error-free PV energy forecast will be used as the optimal solution (lower bound) to evaluate the performance of other EV charging strategies.

Fairness

The studied objective function in Equation (4.1) consists of two terms. The second term is included to minimize the per-unit charging cost paid by each electric vehicle. This term enforces that fairness should be maintained among all EVs in the system. In the presented experiments, fairness is measured by observing the percentage of EVs that satisfy Equation (4.14), and by calculating a fairness index value. Let $[C]$ be the set of per-unit charging costs of each EV. Then, the fairness index $\mathcal{F}[C]$ of this set of EVs is calculated as follows:

$$\mathcal{F}[C] = \frac{1}{1 + \left(\frac{\sigma_{[C]}}{\overline{[C]}}\right)^2} \tag{5.11}$$

where $\sigma_{[C]}$ is the standard deviation of $[C]$, and $\overline{[C]}$ represents the mean value of $[C]$. This fairness index value is in the range $[0, 1]$. If $\sigma_{[C]} = \infty$, then the fairness index will be zero i.e., the studied system would be completely unfair. On the other hand, if $\sigma_{[C]} = 0$, then the fairness index will be unity i.e., the studied system would be completely fair.

Scalability

The desired smart charging algorithm must operate in real-time, and it should be able to handle many electric vehicles. Thus, the applicability of each studied charging strategy can be compared in terms of scalability. The lack of scalability is a significant draw-

back, even if the smart charging strategy is optimal. This is because the optimization time would increase drastically, and the algorithm may not be able to perform optimization when applied to a large-scale smart grid.

Results

Small-scale case study

The small-scale case-study is performed using the distribution network shown in Figure 5.5. The discussed reinforcement learning-based charging strategies are allowed to learn for 30 simulation days i.e., the *learning phase*. The mean rewards (mean of all EVs' average rewards) of the proposed baseline CMAB-based adaptive multi-agent smart charging strategy without any PV estimation and the proposed CMAB-based adaptive multi-agent smart charging strategy with PV estimation are shown in Figure 5.13 and Figure 5.14 respectively. Convergence can be observed within 30 simulation days for any choice of learning strategy (Thompson sampling, UCB, or EXP3). The mean reward $r_{mean}(t)$ of a learning strategy at instant t, with E number of learning EV agents in the system, each observing an instantaneous reward $r_e(t)$ and having an average reward value $r_{avg,e}(t)$, is calculated as follows:

$$r_{mean}(t) = \frac{\sum_{e=1}^{E} r_{avg,e}(t-1) + (r_e(t) - r_{avg,e}(t-1))}{E} \tag{5.12}$$

It can be seen in Figure 5.13 that the EXP3 learning strategy is performing the best at the start of the *learning phase*. The adversarial EXP3 approach can be well suited to handle the non-stationarity in the choice of super arms of other players (EVs) from one player's (EV's) point of view. The mentioned non-stationarity is expected to be at its highest level at the beginning of the simulation-based experimentation. Thus, the EXP3 learning strategy shows this superior performance initially. However, as each agent learns its respective optimal policy, the non-stationarity in the choice of super arms of other EVs from one EV's perspective decreases. Hence, the Thompson sampling-based learning strategy starts performing better than the EXP3-based learning strategy. The UCB-based learning strategy shows inferior performance in this simulation-based experimentation comparatively, but it still converges to a near-

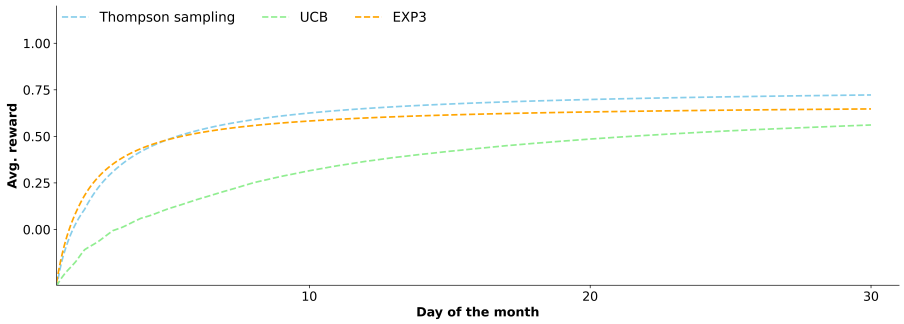

Figure 5.13: Average learning reward of the system when the CMAB-based learning strategy is applied without any PV estimation.

optimal policy. Furthermore, it can be observed in Figure 5.14 that when EV agents are also learning the trend of the daily freely available PV energy production, all learning strategies converge to the same average reward value. This is because EVs also utilize the available PV energy production during the day. Hence, the competition during the low electricity price instants is reduced. This reduced competition favors the UCB and the EXP3 learning strategies, thus achieving the same average reward value in Figure 5.14. It must be noted that the curves shown in Figure 5.14 do not represent the objective function-based cost optimality results of the system. These average reward curves are based on virtual reward values designed to assist the learning of each EV agent. Thus, these curves signify the learning convergence of agents here. The calculation of daily charging costs results to study optimality is made in the upcoming sub-section.

After the *learning phase* is complete, the next 30 simulation days are used for performance evaluation i.e., the *evaluation phase*. Both the uncontrolled charging strategy and the presented MILP optimization-based charging strategy are also evaluated during this *evaluation phase* to draw comparisons.

Constraints satisfaction

The distributions of the voltage are the last bus of the distribution network, and the electrical current (flowing through the electrical line connecting the sub-district SD1 with the district D1 in Figure 5.5), during the *evaluation phase*, is presented in Figure 5.15. These distributions are obtained when the MILP optimization charging strategy is followed. It can be seen that the bus voltage remains within the desired limits of 0.95 pu and 1.05 pu. Additionally, it can be observed that there is no current constraint violation (i.e., no electrical current congestion) in the system if the MILP optimization charging strategy is followed. The obtained results are intuitive, as both of these congestion constraints are modeled as hard constraints in the presented MILP formulation, in Section 5.2.

However, the same cannot be said for the uncontrolled charging strategy. In Figure 5.16, the distributions of the bus voltage and the electrical line current during the *evaluation phase* are shown, when the uncontrolled EV charging strategy is followed. Both of the studied congestion constraints are violated here. There is an under-voltage issue in the network as well as electrical line congestion. During the *evaluation phase*, the distribution network suffers from the shown under-voltage issue for 6.11% of the

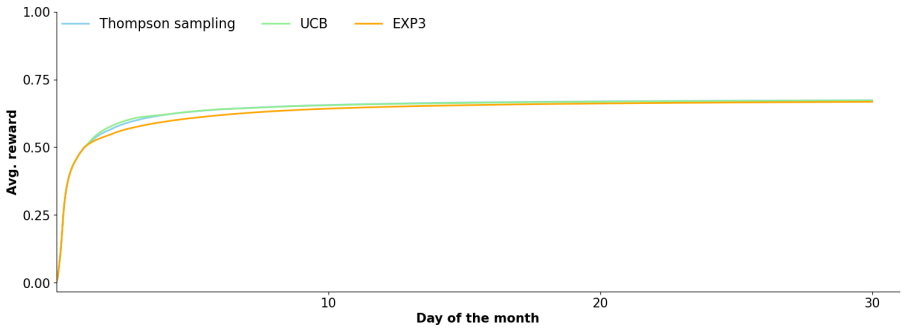

Figure 5.14: Average learning reward of the system when the CMAB-based learning strategy is applied with the PV estimation.

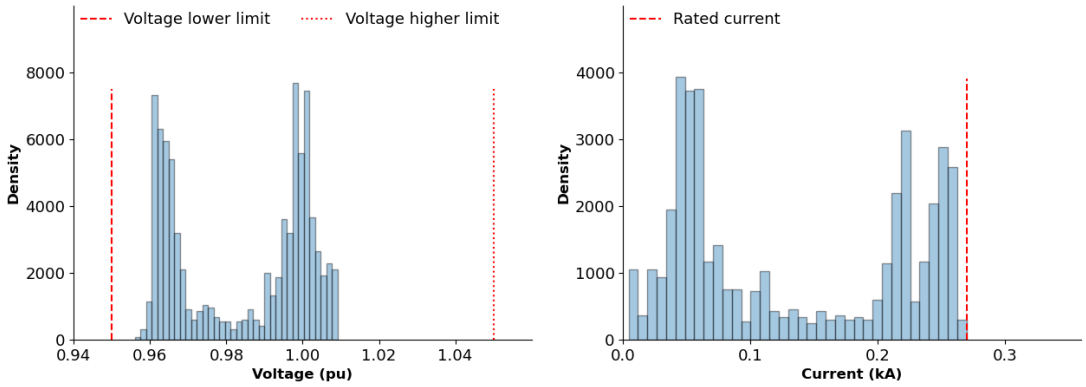

Figure 5.15: Distribution of the bus voltage (left) and the line current (right) obtained through the MILP optimization.

total time (i.e., 30 days) when the uncontrolled charging strategy is followed. Additionally, the distribution network is also congested (electrical line congestion) for 10.14% of the total *evaluation phase*, when EVs are charging without any control. The charging policies of three of three randomly selected electric vehicles during one of the evaluation days are given in Appendix B.

The surface plots of the bus voltages (at the last bus of the studied small-scale distribution network) obtained through the CMAB-based adaptive multi-agent charging strategies (both with and without PV estimation) are shown in Figure 5.17. In both of the shown surface plots, voltage constraint violations at the initial stage of the *learning phase* can be observed. As the *learning phase* progresses, EV agents try to prevent this voltage congestion. Moreover, the magnitude of the under-voltage issue is higher when EV agents are not learning the PV energy production trend compared to when EV agents are learning the daily PV energy production trend. The distributions of the voltage on the last bus of the distribution network, and the electrical current (flowing through the electrical line connecting the sub-district SD1 with the district D1), during the *evaluation phase*, are presented in Figure 5.18 and Figure 5.19.

These distributions are obtained when the CMAB-based adaptive multi-agent charging strategy is followed without any PV estimation, and with each of the mentioned

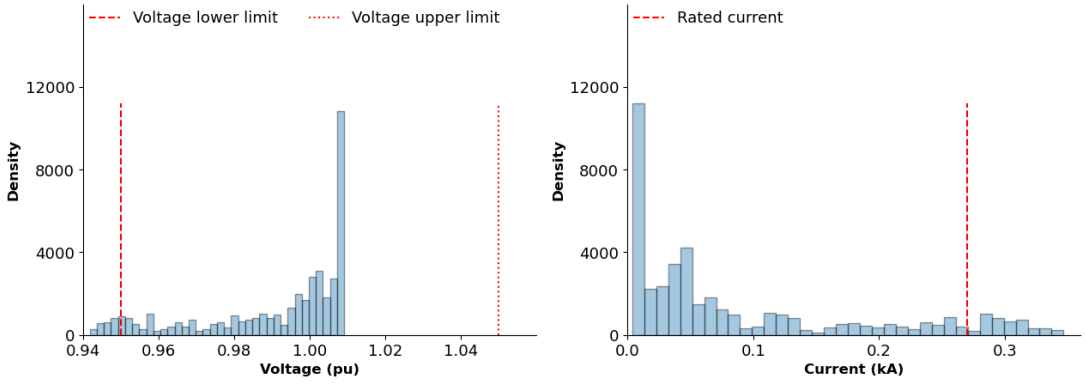

Figure 5.16: Distribution of the bus voltage (left) and the line current (right) obtained through the uncontrolled EVs charging.

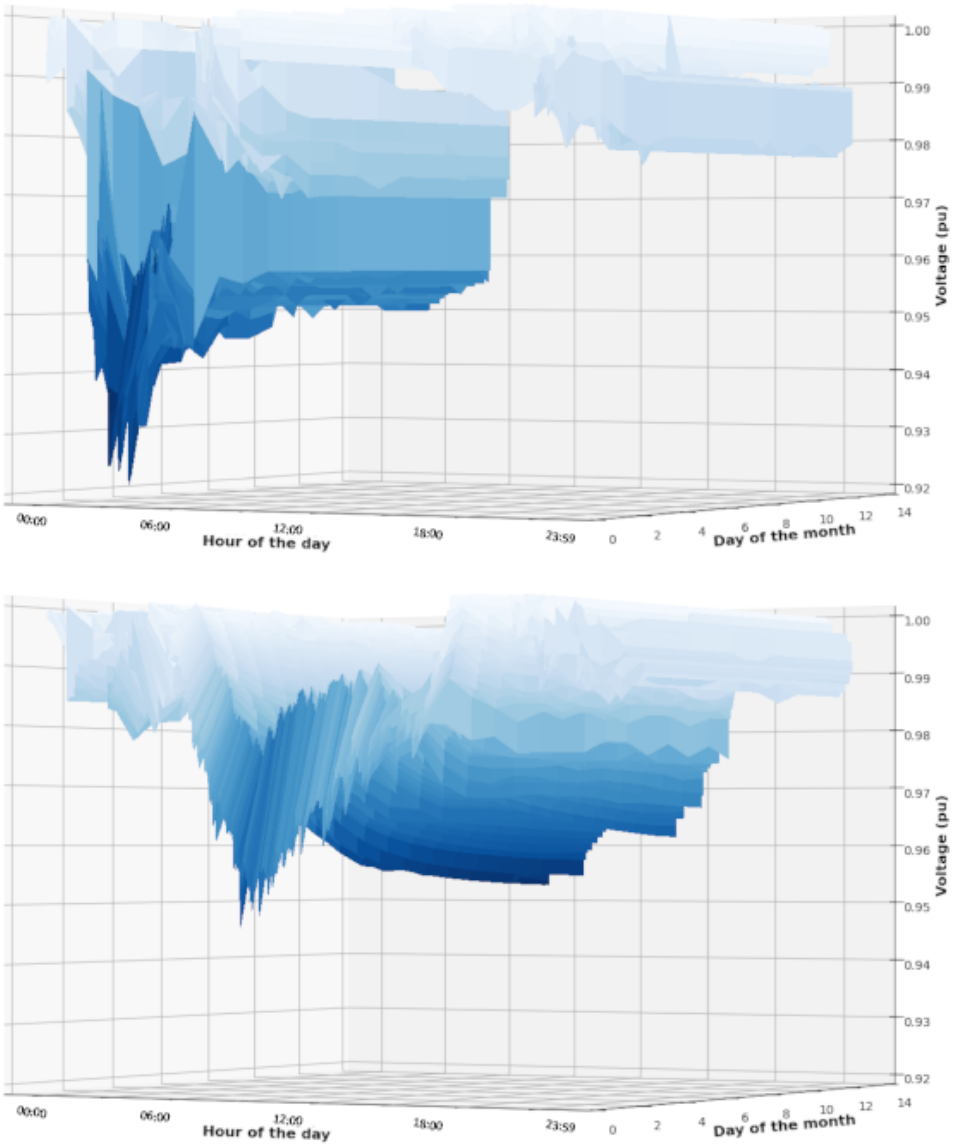

Figure 5.17: Voltage at the last bus when the proposed CMAB-based adaptive multi-agent charging strategy (with Thompson sampling) is followed without any PV estimation (top) and with PV estimation (bottom).

learning strategies (Thompson sampling, UCB, and EXP3). It can be seen that both congestion constraints (voltage and current) are satisfied by all of the learning strategies. These constraints are also satisfied when EV agents follow the proposed CMAB-based adaptive multi-agent smart charging with PV estimation. The resulting distributions are shown in Figure 5.20 and Figure 5.21. Here, it can also be observed that no constraints are violated when our proposed adaptive multi-agent charging strategy is followed. Thus, in terms of constraint satisfaction, all of the studied EV charging strategies are working as desired, except for the uncontrolled EV charging strategy. This unsought functionality of the uncontrolled EV charging strategy is clearly visible in Figure 5.22. In Figure 5.22, the bus voltage (at the last bus) and the line current (through the electrical line connecting the sub-district SD1 with the district D1 in Fig-

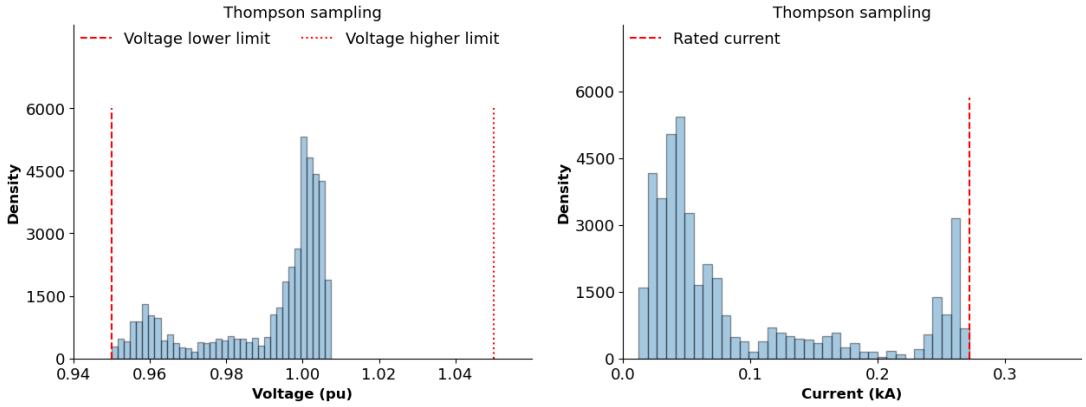

Figure 5.18: Distribution of the bus voltages (left) and the line currents (right) obtained through the CMAB-based adaptive multi-agent charging strategy without any PV estimation (with Thompson sampling).

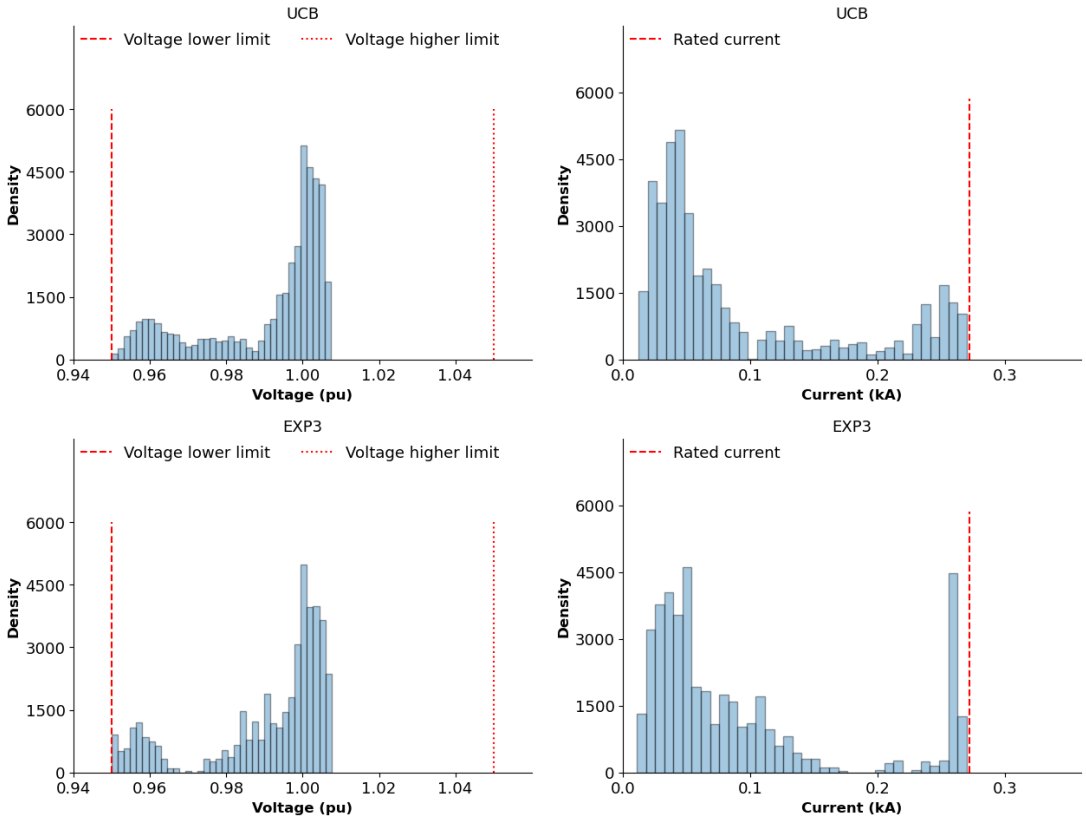

Figure 5.19: Distribution of the bus voltages (left) and the line currents (right) obtained through the CMAB-based adaptive multi-agent charging strategy without any PV estimation (with UCB and EXP3).

ure 5.5) are shown, for all of the studied EV charging strategies. The shown curves correspond to one of the days during the *evaluation phase*. A peak load demand in the evening causes congestion in the studied electrical distribution network.

Other mentioned EV smart charging strategies manage this peak load demand, as shown in Figure 5.22. The CMAB-based smart charging strategies (without any PV

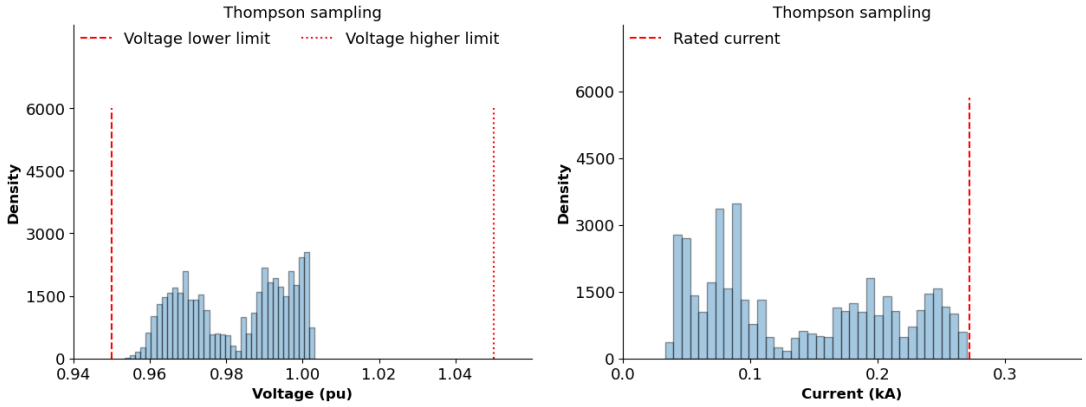

Figure 5.20: Distribution of the bus voltages (left) and the line currents (right) obtained through the CMAB-based adaptive multi-agent charging strategy with the PV estimation (with Thompson sampling).

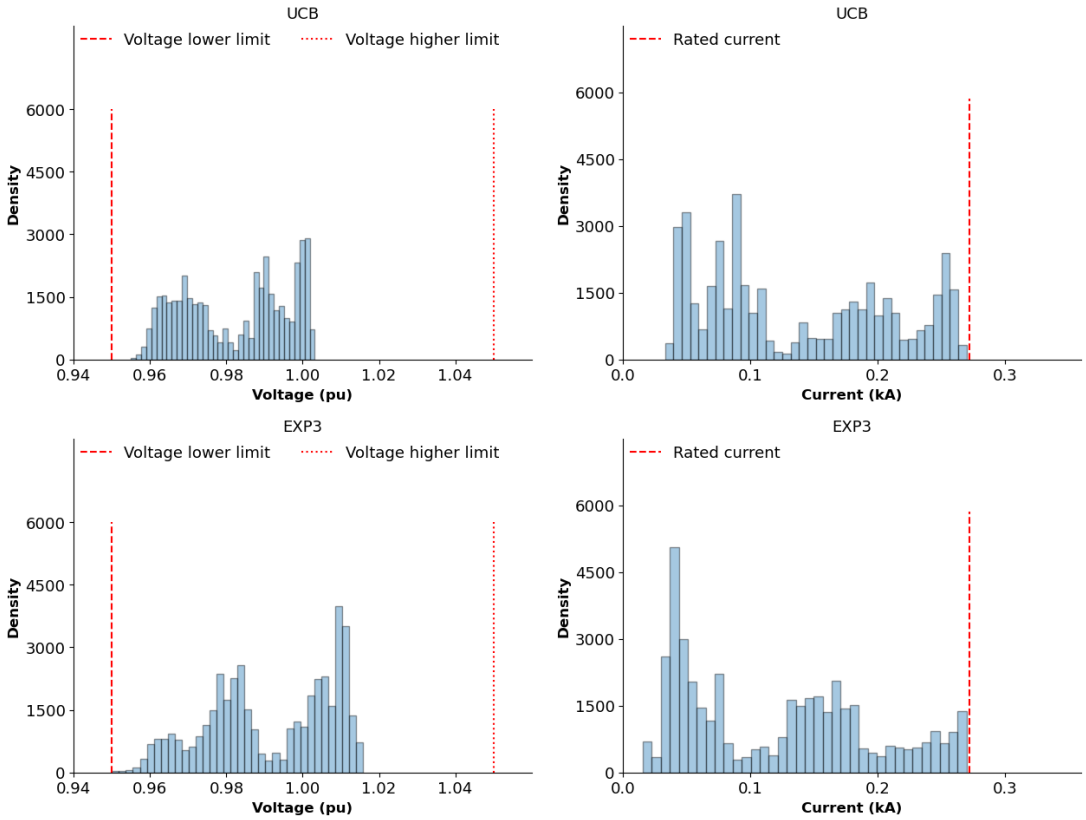

Figure 5.21: Distribution of the bus voltages (left) and the line currents (right) obtained through the CMAB-based adaptive multi-agent charging strategy with the PV estimation (with UCB and EXP3).

estimation) manage the peak load demand by shifting the load during the early hours of the day (when the electricity price is not expensive) while ensuring that the distribution network does not get congested. This shift in EV load would benefit both prosumers and DSOs. However, this mentioned strategy may not be optimal as it does not learn and utilizes the freely available PV energy production. The MILP optimization smart

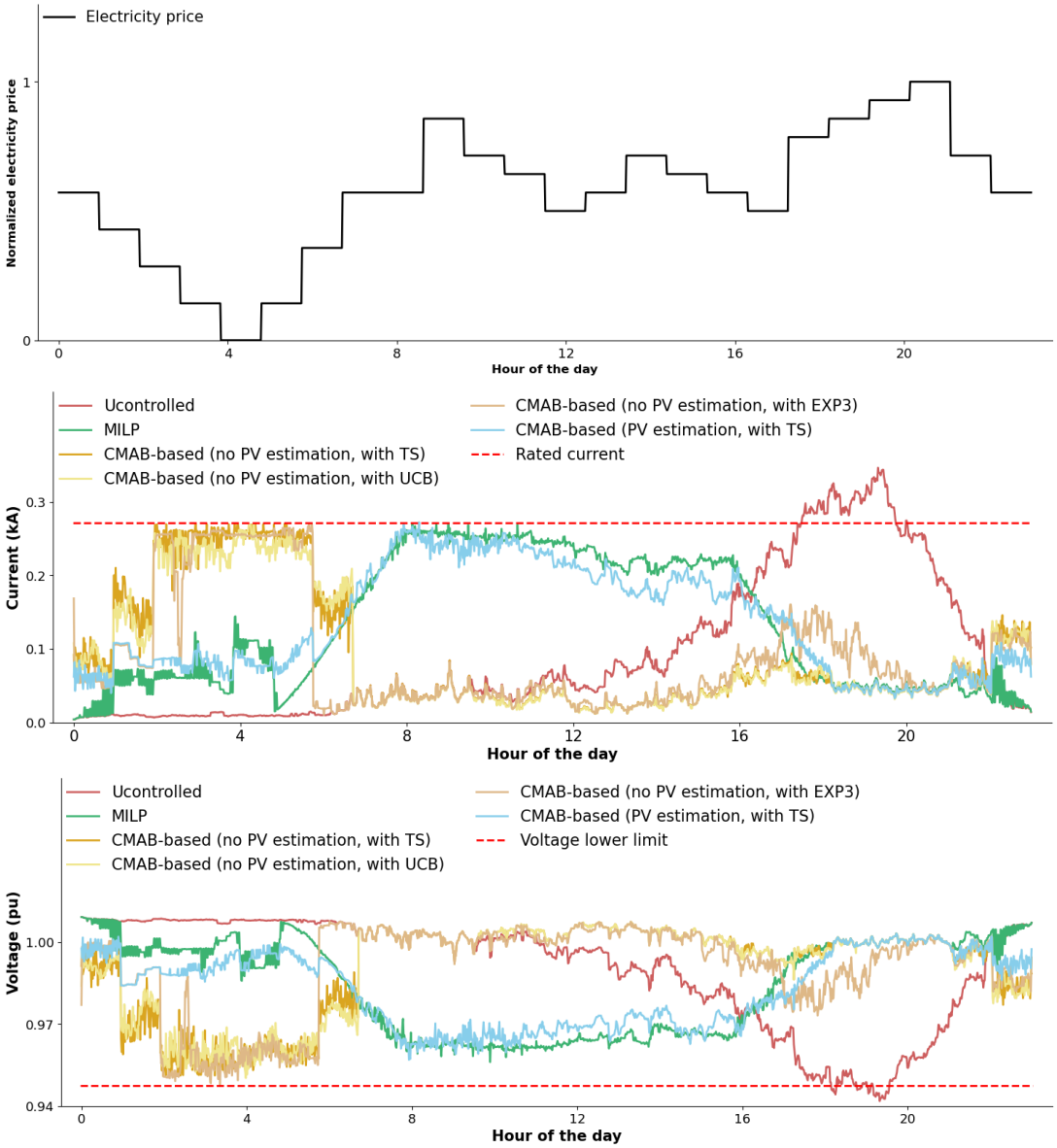

Figure 5.22: Daily electricity price (top). Electrical line currents (middle) and voltages (bottom) comparison during a single day.

charging strategy produces the optimal solution, as zero PV forecast error while performing MILP optimization has been considered here. Thus, the MILP optimization smart charging strategy will utilize all of this accurately known PV energy to minimize the charging costs of each EV. On the other hand, our proposed CMAB-based adaptive multi-agent system learns the trend of this daily PV energy production and does not require this time series data as an input. Yet, it manages to produce a near-optimal (close to the optimal MILP solution) solution. This is a significant contribution of the proposed adaptive multi-agent multi-armed bandit smart charging system as it produces near-optimal solutions while considering real-life uncertainties.

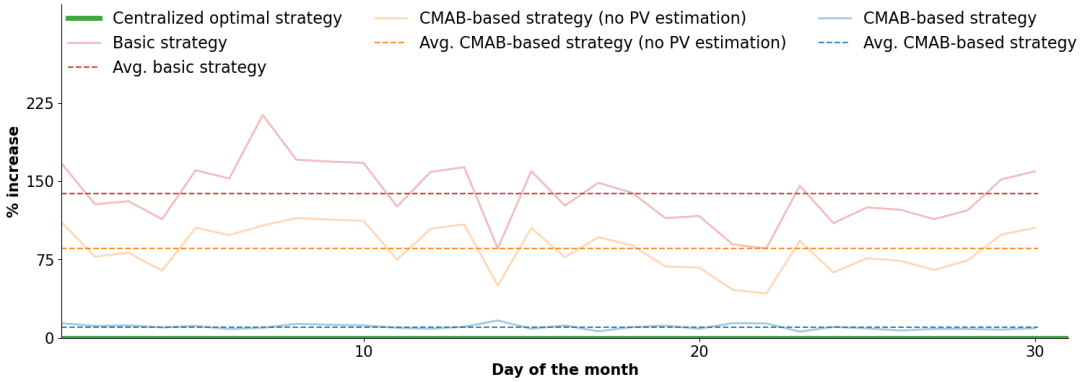

Figure 5.23: Optimality gap comparison compared to the centralized MILP lower bound.

Optimality

As discussed earlier, the MILP optimization strategy with no PV forecast error is considered the optimal charging strategy. Thus, the optimality of each charging strategy is evaluated by calculating the percentage increase in the total daily charging cost compared to the optimal MILP total daily charging cost. The obtained optimality results are presented in Figure 5.23. The horizontal axis is considered the optimal MILP daily charging cost, while the vertical axis represents the percentage increase compared to the optimal horizontal axis. It can be seen in Figure 5.23, that the uncontrolled EV charging strategy holds the highest optimality gap. The average percentage increase when the uncontrolled EV charging strategy is followed during the *evaluation phase* is 138.03%. This increment in the total daily charging cost is expected as the peak EV load demand is in the evening when the electricity price is at its highest, as shown in Figure 5.22.

The total daily charging cost can be reduced if the CMAB-based adaptive multi-agent smart charging without any PV estimation is adapted. In this charging strategy, each EV agent minimizes its daily charging cost by charging when the electricity price is low, without causing any congestion in the distribution network. However, this charging strategy may still be far from optimal as the freely available PV energy production has not been utilized by EV agents during the day. This is evident in Figure 5.23. The average percentage increase, when this EV charging strategy is followed during the *evaluation phase*, is 85.55%. The obtained average optimality gap results for each learning strategy are summarized in Table 5.3, for both the CMAB-based learning strategies with and without the PV estimation.

Finally, the proposed CMAB-based adaptive multi-agent smart charging strategy with the PV estimation (with Thompson sampling learning strategy) is closest to the optimal MILP optimization charging strategy, shown in Figure 5.23. As discussed earlier in Figure 5.22, EVs make an estimate of the daily PV energy production and utilize it to charge during the day without paying any cost. This strategy, as a result, further minimizes the total daily charging cost of EVs. The average percentage increase during the *evaluation phase* is 10.01%, if the proposed adaptive multi-agent EV smart charging strategy is followed. The choice of learning strategy can have a minor impact on this average percentage increment compared to the optimal MILP.

CMAB-based smart charging (without any PV estimation)

Learning strategy	Optimality gap (%)
Thompson sampling	85.55
UCB	109.62
EXP3	97.12

CMAB-based smart charging (with PV estimation)

Learning strategy	Optimality gap (%)
Thompson sampling	10.02
UCB	10.02
EXP3	10.04

Table 5.3: Selected learning strategy impact on the optimality gap.

Centralized MILP optimization charging strategy

% of EVs with $SoC_{e,depart}$	Fairness index
100	1

CMAB-based smart charging (without PV estimation)

Learning strategy	% of EVs with $SoC_{e,depart}$	Fairness index
Thompson sampling	100	0.99
UCB	100	0.99
EXP3	100	0.99

CMAB-based smart charging (with PV estimation)

Learning strategy	% of EVs with $SoC_{e,depart}$	Fairness index
Thompson sampling	100	0.99
UCB	100	0.99
EXP3	100	0.99

Table 5.4: Fairness comparison for the small-scale case study.

This impact occurs because each learning strategy converges to the same mean reward value, as shown in Figure 5.14. The impact of the selection of the learning strategy on the optimality gap for the proposed smart charging strategy is also listed in Table 5.3.

Fairness

Fairness should be maintained among all EV agents in an ideal decentralized system. The fairness comparison between the MILP optimization charging strategy and the CMAB-based adaptive multi-agent smart charging strategies is presented in Table 5.4. The uncontrolled EV charging strategy has not been included in the comparison as it is not an optimization-based charging strategy. In the case of MILP optimization, as it is a centralized optimization approach with the fairness term included in the objective function, it is entirely fair. All EVs attain their desired $SoC_{e,depart}$ at their respective departure times. This constraint is also satisfied by both of the CMAB-based smart charging strategies. The fairness index value of the MILP optimization strategy, calculated using Equation (5.11), is equal to 1. Moreover, these values are sufficiently close to 1 for the proposed CMAB-based charging strategies. It confirms that fairness among all EV agents is taken into account by the proposed adaptive multi-agent multi-armed bandit smart charging system.

Scalability

All of the studied EV charging strategies are scalable up to this point. The solutions obtained through both the centralized MILP and the CMAB-based adaptive multi-agent EV charging strategies have been discussed. However, there were only 55 EVs in the studied small-scale distribution network. Thus, a large-scale case study with 10,175 EVs is proposed to evaluate this performance metric better. This large-scale case study is discussed next.

Large-scale case study

The large-scale simulation study is performed using the electrical distribution network model given in Figure 5.6. There are a total of 10,175 intelligent EV agents in the studied large-scale system. The centralized MILP optimization algorithm is unable to perform optimization here due to a large number of agents in the system (resulting in extremely long computing times and memory requirements). However, proposed CMAB-based adaptive multi-agent strategies still manage to converge. Convergence is also observed within 30 simulation days of training, i.e., the *learning phase*. This confirms statistically that the proposed CMAB-based adaptive multi-agent system is scalable in terms of convergence to a good solution (as the convergence time has not increased here compared to the small-scale study). The plots of mean rewards (moving average of all EV's mean reward values) of both CMAB-based strategies without and with PV estimation are shown in Figure 5.24 and Figure 5.25. Similar performance trends can be observed here as well compared to the small-scale studies, i.e., Thompson sampling outperforms UCB and EXP3 when no PV estimation is done, and all learning strategies are converging to approximately the same average reward value when PV estimation is performed.

Once the *learning phase* is complete, the next 30 simulation days are used to evaluate the performance of the learning strategies, i.e., the *evaluation phase*. The performance is evaluated here in terms of improvement compared to the uncontrolled charging strategy (and not in terms of the optimality gap). This is because the cen-

tralized MILP solution can not be obtained for the studied large-scale system and thus the optimal solution is unknown here. Nonetheless, it has already been shown in the small-scale simulation study section that the proposed CMAB-based adaptive multi-agent smart charging system with PV estimation is capable of producing near-optimal solutions.

Constraints satisfaction

Ideally, the deployed charging strategy should satisfy both grid and prosumer constraints. The voltage (at the last bus of the sub-district SD1 in Figure 5.6) distribution and electrical current (flowing through the grid transformer line in Figure 5.6) distribution, if uncontrolled charging strategy is followed, is shown in Figure 5.26. The shown distributions are obtained based on the results during the *evaluation phase*. It can be observed that both electrical current and voltage constraints are violated if the uncontrolled charging strategy is followed. This is due to a large number of EVs charging simultaneously without any control. The studied electrical distribution network suffers from an under-voltage issue for 4.93% of the total evaluation period (30 simulation days). It is also suffering from electrical line congestion for 15.48% of the total evaluation period when the uncontrolled charging strategy is applied.

The distribution results obtained when the CMAB-based smart charging strategy (without PV estimation) is followed are shown in Figure 5.27 and Figure 5.28. These distributions are obtained using all three of the aforementioned learning strategies (Thompson sampling, UCB, and EXP3) combined with the CMAB-based adaptive multi-agent charging strategy without the use of any PV estimation. It can be observed that grid constraints are not violated in all of the studied learning strategies. The obtained distributions when PV estimation is performed by the CMAB-based adaptive multi-agent smart charging system combined with the Thompson Sampling learning strategy are shown in Figure 5.29. It is evident that grid constraints are also satisfied in this mentioned case. It should be noted that the shown voltage results are for the voltage at the last bus of the sub-district SD1 in Figure 5.6, and the plotted line results correspond to the electrical current flowing through the grid transformer line in Figure 5.6.

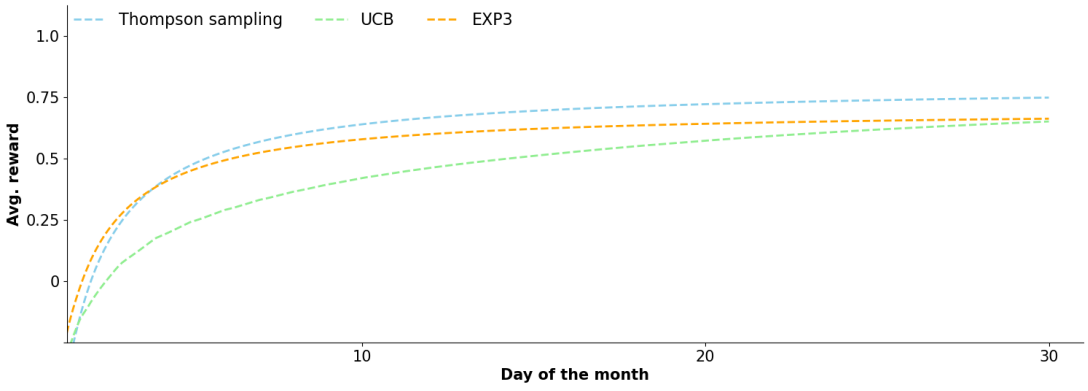

Figure 5.24: Average learning reward of the system when the CMAB-based learning strategy is applied without any PV estimation.

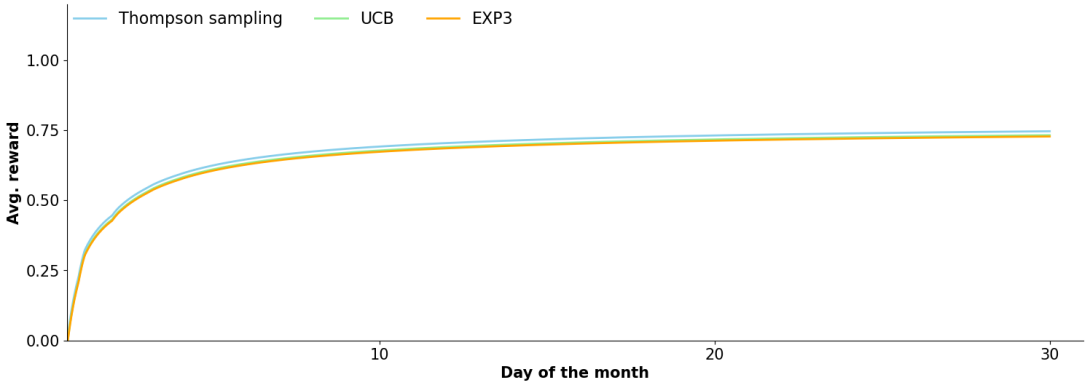

Figure 5.25: Average learning reward of the system when the CMAB-based learning strategy is applied with the PV estimation.

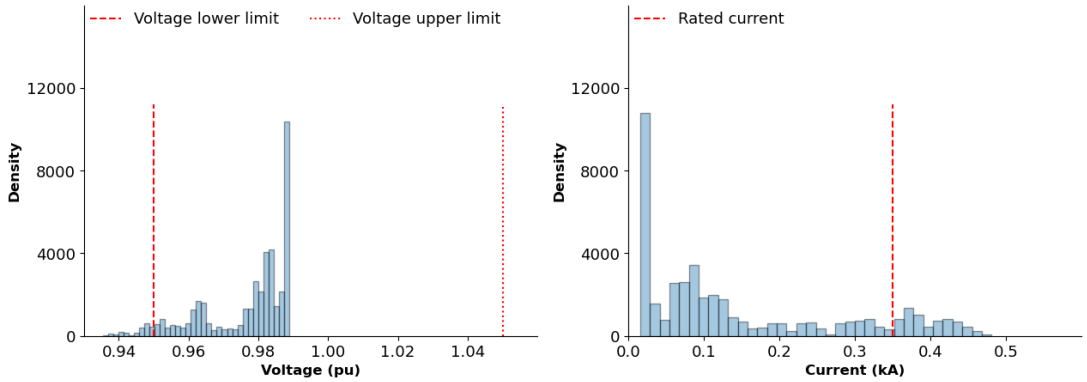

Figure 5.26: Distribution of the bus voltage (left) and the line current (right) obtained through the uncontrolled EVs charging.

Daily current and voltage profiles for the studied charging strategies during one of the evaluation days are shown in Figure 5.30. It can be verified that a peak load demand will occur during the evening if the uncontrolled charging strategy is followed.

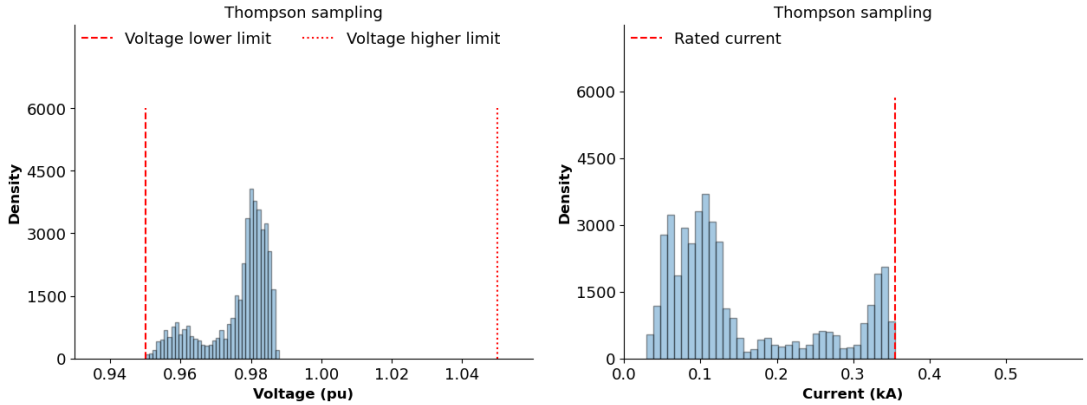

Figure 5.27: Distribution of the bus voltages (left) and the line currents (right) obtained through the CMAB-based adaptive multi-agent charging strategy without any PV estimation (with Thompson sampling).

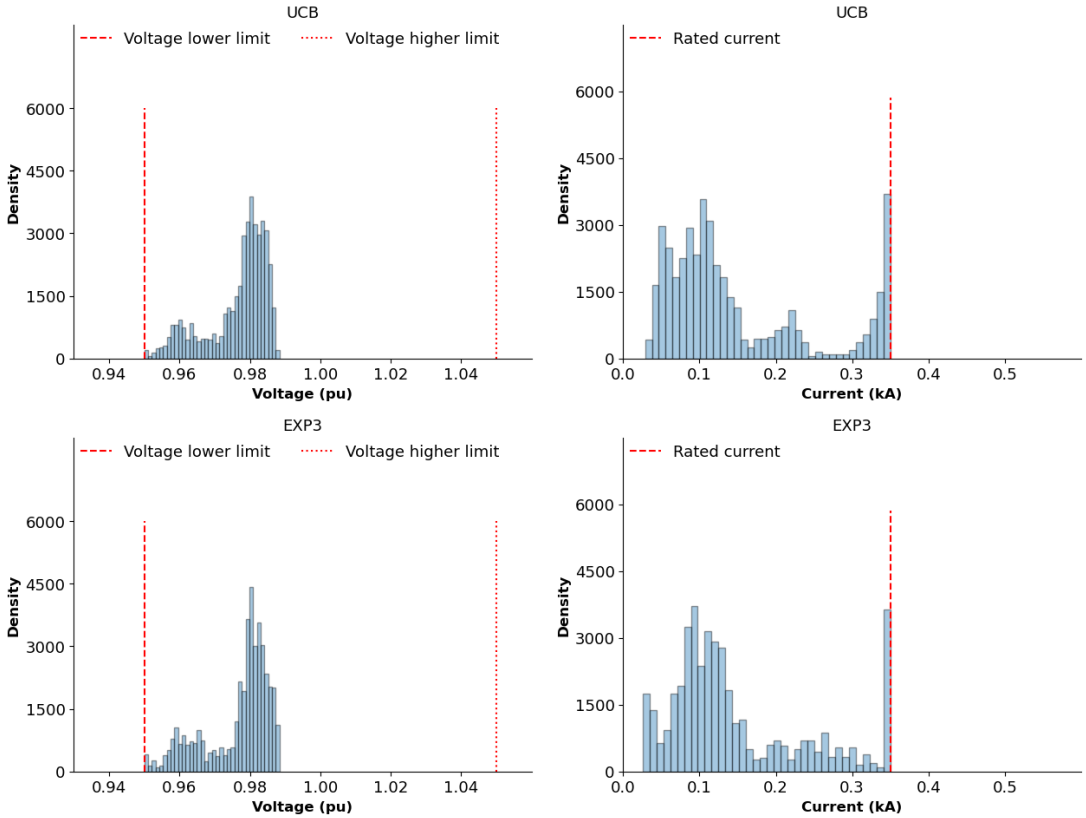

Figure 5.28: Distribution of the bus voltages (left) and the line currents (right) obtained through the CMAB-based adaptive multi-agent charging strategy without any PV estimation (with UCB and EXP3).

This is due to a high number of EV owners returning home and plugging in their EVs to charge. This peak load demand can be avoided if any of the proposed CMAB-based adaptive multi-agent smart charging strategies are utilized. Furthermore, it can be verified in Figure 5.30 that EVs also utilize the freely available PV energy by charging

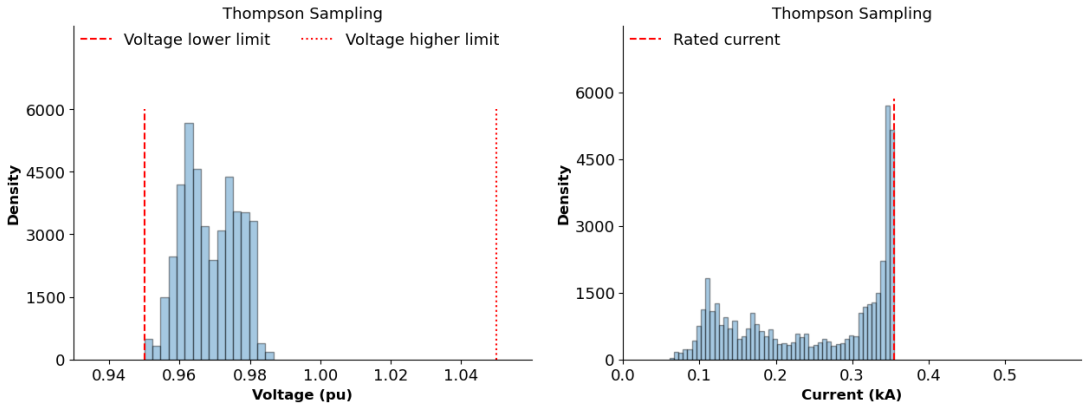

Figure 5.29: Distribution of the bus voltages (left) and the line currents (right) obtained through the CMAB-based adaptive multi-agent charging strategy with the PV estimation (with Thompson sampling).

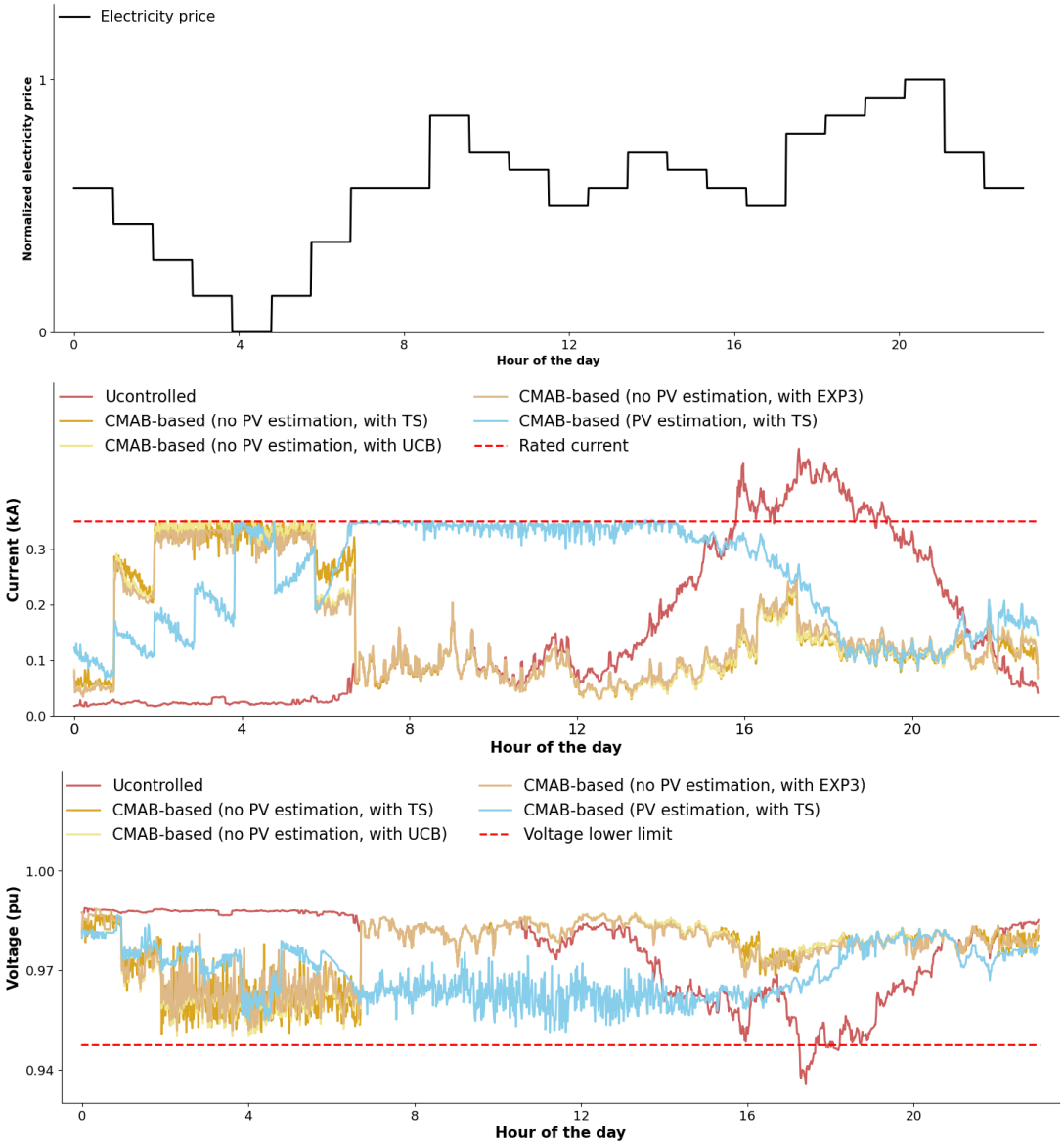

Figure 5.30: Daily electricity price (top). Electrical line currents (middle) and voltages (bottom) comparison during a single day.

during the daytime when PV estimation is combined with the proposed CMAB learning algorithm. This will have a significant impact on the daily charging cost reduction of each EV. It can be concluded that the uncontrolled charging strategy fails to satisfy grid constraints when applied to a large-scale electrical distribution network. Whereas, the proposed CMAB-based adaptive multi-agent strategy manages to satisfy grid constraints irrespective of the utilized learning strategy and whether the PV estimation is made or not.

Cost reduction

The centralized MILP strategy is not able to perform optimization when applied to the studied large-scale system. Thus, the theoretical optimal solution remains unknown

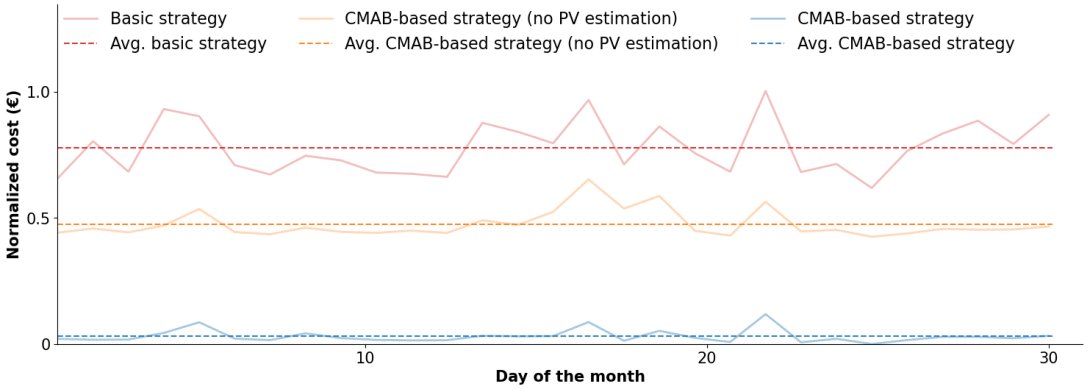

Figure 5.31: Cost reduction comparison compared to the uncontrolled strategy upper bound.

here. However, the quality of any obtained solution can also be evaluated by comparing it with the solution obtained through the uncontrolled strategy. The uncontrolled charging strategy solution is considered as the upper bound here on the cost. Thus, the comparison is made here in terms of cost reduction achieved by any given strategy compared to the cost of the uncontrolled charging strategy. The obtained cost reduction results during the *evaluation phase* are plotted in Figure 5.31.

It can be seen in Figure 5.31 that the proposed CMAB-based adaptive multi-agent smart charging strategy is performing better than the uncontrolled charging strategy (with Thompson Sampling) even without any PV estimation. An average cost reduction of 23.21% is observed compared to the uncontrolled charging strategy. However, this daily charging cost can be further reduced if the CMAB-based adaptive multi-agent system also utilizes an estimation regarding the instantaneous PV energy production. This is also evident in Figure 5.31 as the average cost reduction achieved by the CMAB-base smart charging strategy with PV estimation is 77.58%. Thus, it can be deduced that although optimal solution (lower bound) remains unknown here. However, the improvement can still be confirmed by comparing the solutions obtained through the CMAB-based strategies with the basic charging strategy (upper bound). The choice of learning strategy can have an impact on the performance of the CMAB-based system as well. This impact on the daily charging cost is summarized in Table 5.5. It can be observed that the impact of the choice of the learning strategy is little when no PV estimation is made, and it is insignificant when PV estimation is made by the proposed adaptive multi-agent system.

Fairness

The fairness comparison is presented in Table 5.6. The fairness comparison parameters considered here are the same as they were in the small-scale study subsection, i.e., the % of EVs with the desired SoC at their departure time, and the fairness index. The fairness index remains undefined for the uncontrolled strategy here as well. This is because the uncontrolled charging strategy is not an optimization methodology and thus the EVs remain uncontrolled. The fairness index is calculated only for the optimization control strategies here. The % of EVs managing to achieve the desired SoC is 100% for all the charging strategies compared in this section. The fairness index comes

CMAB-based smart charging (without any PV estimation)

Learning strategy	Cost reduction (%)
Thompson sampling	23.21
UCB	22.13
EXP3	22.10

CMAB-based smart charging (with PV estimation)

Learning strategy	Cost reduction (%)
Thompson sampling	77.58
UCB	77.52
EXP3	77.51

Table 5.5: Selected learning strategy impact on the cost reduction.

out to be around 0.99 for the proposed CMAB-based smart charging strategy (for all learning strategies, and with/without PV estimation). This confirms numerically that the proposed CMAB-based is indeed taking fairness into account among different EV agents, and thus the system is converging to a solution that maintains fairness in the system.

Scalability

In the studied problem, scalability is an important factor as the control algorithm is expected to optimize large-scale practical electrical distribution networks. The centralized MILP optimization does not manage to optimize the studied large-scale smart grid due to its lack of scalability. As it belongs to the NP (non-deterministic polynomial) class of problems, its optimization time increases significantly as the number of agents is increased in the system. This has also been observed earlier in Figure 3.11. On the other hand, the proposed CMAB-based adaptive multi-agent system can be considered scalable as it managed to control the studied large-scale system efficiently. As discussed earlier, the computation time of each agent (i.e., time to find the estimated super arm to play) is independent of the total number of agents in the system. Thus, the decentralization of decision-making in smart grids has evidently helped in designing a scalable system that can operate on large-scale smart grids.

5.4 Conclusion

This chapter commenced by transforming the multi-agent multi-armed bandit system for smart charging, as presented in the previous chapter, into an adaptive multi-agent multi-armed bandit system. This transformation involved leveraging the concepts of adaptive multi-agent system theory discussed in Chapter 2. The proposed system has been compared with a number of baseline charging strategies for performance evalua-

Uncontrolled charging strategy

% of EVs with $SoC_{e,depart}$	Fairness index
100	-

CMAB-based smart charging (without PV estimation)

Learning strategy	% of EVs with $SoC_{e,depart}$	Fairness index
Thompson sampling	100	0.99
UCB	100	0.99
EXP3	100	0.99

CMAB-based smart charging (with PV estimation)

Learning strategy	% of EVs with $SoC_{e,depart}$	Fairness index
Thompson sampling	100	0.99
UCB	100	0.99
EXP3	100	0.99

Table 5.6: Fairness comparison for the large-scale case study.

tion. These baseline strategies include the uncontrolled charging strategy, centralized MILP optimization, and a variation of the proposed CMAB-based adaptive multi-agent strategy but without any PV estimation. The performance of each charging strategy has been measured through case studies and in terms of optimality, scalability, and constraint satisfaction.

Observing the results obtained in Section 5.3, it can be concluded that the uncontrolled charging strategies do not face any concern related to scalability. However, it would result in a far-from-optimal solution along with constraint violations. On the other hand, the centralized MILP optimization approach faces scalability challenges. The proposed CMAB-based adaptive multi-agent system managed to tackle the scalability challenges of the centralized MILP optimization through decentralization of the system, and it managed the real-life uncertainties by using combinatorial multi-armed bandit learning. This resulted in a system that is real-time, scalable, satisfies all required constraints, and is near-optimal. It was also observed that if PV energy generation estimation is done in the proposed system then the optimality of the system is improved significantly. This proposed system also maintains fairness among electric vehicles in the system and ensures the data privacy of each EV owner's personal data. The proposed adaptive multi-agent multi-armed bandit system can also be adapted to control other flexible entities of a smart grid to optimize its energy flows.

Conclusions and perspectives

Work summary

This thesis primarily focused on the design of a smart grid control system aimed at optimizing energy flows within smart grids. The introductory chapter (Chapter 1) provided a comprehensive overview of potential challenges arising from the integration of novel technologies like renewable energy sources and distributed energy generation in power systems. These challenges, primarily linked to uncertainty, highlighted the significance of smart grid control solutions in facilitating the seamless integration of novel grid elements while maintaining grid stability. In this chapter, a number of existing smart grid control solutions were presented and classified according to their system architecture, namely centralized, hierarchical, and decentralized approaches. The analysis emphasized the benefits of decentralization in terms of scalability and real-time operational capabilities, which represent the core contributions of this research work.

Chapters 2 and 3 presented the first of the two proposed decentralized smart grid control systems, leveraging adaptive multi-agent system theory to maintain decentralization and scalability. This system was designed to address the problem of real-time grid balancing by controlling electric vehicle instantaneous charging and discharging. Different smart grid elements were modeled as software agents, each with distinct goals. These agents were collaboratively achieving system functionality. Electric vehicle agents, the decision-making entities, employed a heuristic process for decision-making. Through deterministic and pseudo-stochastic simulation-based experiments (Chapter 3), the adaptive multi-agent system demonstrated real-time, scalable, and near-optimal performance. Although it was concluded that anticipative capabilities could further enhance its efficiency.

In Chapters 4 and 5, the focus shifted to incorporating anticipative capabilities through combinatorial multi-armed bandit theory. A comprehensive introduction to multi-armed bandit and combinatorial multi-armed bandit was presented in Chapter 4. Their faster convergence in comparison to conventional algorithms, like deep Q-learning, rendered them a compelling reinforcement learning approach for integrating anticipative capabilities into the system agents. A multi-agent multi-armed bandit system was introduced in Chapter 4, addressing smart electric vehicle charging under uncertainties in daily photovoltaic energy production. Building on this, the final adaptive multi-agent multi-armed bandit system (Chapter 5) tackled the same smart charging problem, exhibiting near-optimal performance, scalability, real-time responsiveness, and fairness. Furthermore, it was argued that the proposed system's adaptability allows for its application to other flexible entities in smart grids as well.

Conclusions

This section presents the key findings of the research carried out in this thesis manuscript. Conclusions presented here aim to verify the veracity of the hypotheses put forth in

Chapter 1 based on the work presented in all of the earlier chapters. At the end of this section, a comprehensive summary of the two decentralized smart grid control systems proposed in this thesis, which constitute the primary contributions of this research, is also provided. The very first hypothesis made in Chapter 1 is as follows:

> *The theory of adaptive multi-agent systems can serve as a valuable framework for developing an effective real-time decentralized energy management control system for large-scale active electrical distribution networks.*

The study aimed to examine the feasibility of developing a decentralized control system using adaptive multi-agent systems (AMAS) to manage a large-scale smart grid in real-time. To test the above-mentioned hypothesis, an AMAS was designed and evaluated in Chapters 2 and 3. The system's objective was to optimize grid balancing by effectively utilizing flexible electric vehicles within the distribution network. Simulation-based experiments compared the performance of the proposed AMAS with two baseline strategies, namely the uncontrolled strategy and the centralized MILP optimization strategy. Results from Section 3.2 indicated that the proposed system achieved near-optimal solutions, demonstrating its efficient optimization capability. Furthermore, as seen in Figure 3.11, the proposed adaptive multi-agent control system is expected to be scalable when implemented in a real-world smart grid, as opposed to a centralized system, which may experience bottlenecks due to its inability to scale.

This observation extended to the second system presented in Chapter 5, where the adaptive multi-agent systems theory was combined with multi-armed bandit learning to optimize electric vehicle charging while adhering to the constraints of different market actors. Simulation-based experiments in Section 5.3 confirmed that this system achieved near-optimal solutions while maintaining its scalability and real-time capabilities. A key takeaway from these studies is that the adaptive multi-agent systems theory serves as a potent tool for designing scalable and real-time control systems capable of addressing various complex, large-scale, and real-time smart grid optimization challenges. Nonetheless, it is important to emphasize that the optimality of an adaptive multi-agent system highly depends on the design of the decision-making agents within the system. The lack of anticipative capabilities in the agents, as noticed in the design of the reactive heuristic adaptive multi-agent system in this thesis, could affect its efficiency, especially when dealing with uncertainties in smart grid optimization problems. It may lead to challenges in satisfying various constraints required for smooth grid operations. Finally, the cooperation mechanism in adaptive multi-agent systems results in self-organization which makes such systems adaptable to changes. The second hypothesis made in Chapter 1 is stated below:

> *The incorporation of multi-armed bandit class of reinforcement learning algorithms into the aforementioned adaptive multi-agent system can enhance the system's performance under real-life uncertainties, while simultaneously preserving its scalability and real-time operations capabilities.*

Uncertainties play a crucial role in the efficient operation of smart optimal grid control systems. These uncertainties can arise from various sources, such as fluctuations in renewable energy sources' instantaneous energy production. Additionally, in

decentralized multi-agent systems, each agent's actions can introduce uncertainty from the perspective of other agents. Chapter 3 highlighted the significance of anticipative (learning) capabilities in decision-making agents of adaptive multi-agent systems for improved performance under uncertainties. To address these real-life uncertainties, the second hypothesis proposed the use of multi-armed bandit learning algorithms. It was conjectured that these algorithms could maintain the system's scalability and real-time operation capabilities. The validity of this hypothesis was confirmed through the detailed design and simulation-based experiments of the combinatorial multi-armed bandit learning-based adaptive multi-agent system in Chapter 5.

The simulation results depicted in Figure 5.22 showed that the proposed system effectively considers uncertainty associated with other agents' actions in the decentralized system, ensuring compliance with grid stability constraints. Furthermore, as shown in Figure 5.23, the proposed decentralized system substantially reduced the system's cost by employing PV energy generation estimations to manage uncertainty in PV instantaneous energy data. The proposed system exhibits convergence to a solution with an optimality gap of 10.04%, as evidenced by Table 5.3. It was suggested that this optimality gap could potentially be further reduced by incorporating advanced photovoltaic forecasting techniques. Nevertheless, it is crucial to acknowledge that achieving a null optimality gap is unattainable for the studied smart grid problem due to the inherent impossibility of obtaining an instantaneous photovoltaic energy generation forecast without any error. Chapter 5 also established that the proposed combinatorial multi-armed algorithm maintains the system's scalability and real-time operation capabilities. The system is capable of finding its estimated optimal policy at any given instant in $O(m)$ where m is the total number of decision instants in the studied smart grid combinatorial optimization problem.

The main lesson here is that the application of the multi-armed bandit class of reinforcement learning algorithms in smart grid control is a promising approach, particularly for managing real-life uncertainties in a decentralized manner. These algorithms exhibit lower computational time and memory requirements compared to other commonly used reinforcement learning algorithms like deep-Q learning, resulting in improved scalability and faster convergence. The potential advantages of multi-armed bandit algorithms can be harnessed to address various smart grid optimization challenges beyond those studied in this thesis.

To summarize, two adaptive multi-agent systems have been discussed in-depth in this manuscript to optimize energy flows in smart grids. First, the system proposed in Chapter 2 relies solely on heuristics to perform optimization. This system was shown to have performance degradation under uncertainties due to a lack of anticipative abilities in the system. Finally, combinatorial multi-armed bandit learning was utilized to enable agents with anticipative capabilities in Chapter 4. Combinatorial multi-armed bandit learning helped in tackling stochasticities in the system and thus made the system more optimal under uncertain conditions. Faster convergence of these combinatorial multi-armed bandit learning algorithms compared to more commonly used reinforcement learning algorithms like DQN learning is a significant advantage, especially for smart grid control applications when a perfect oracle is not available and faster convergence can bring more economic advantage. The final proposed system in

Section 5.1 is:

- **Decentralized:** It is fully decentralized. There is no central decision-making entity, rather each agent in the system is making decisions for itself.
- **Real-time:** It is able to perform control operations in real-time (in seconds or minutes).
- **Scalable:** It can scale and operate on large-scale smart grids.
- **Near-optimal:** It gives near-optimal solutions even under stochastic conditions.
- **Tackling stochasticities:** It takes into account different stochasticities that may exist in a practical decentralized smart grid control system.
- **Fair:** It maintains fairness among different decision-making agents present in the system.
- **Model-free:** It does not require any model of the electrical grid for its desired operation.
- **Adaptable:** It can be used to control various flexible elements that might be present in a smart grid and is adaptable to changes.

Perspectives

This section discusses several novel research avenues that have emerged as a result of the work presented in this thesis. These research perspectives cover a range of complexities, spanning from short-term to medium-term exploration.

Potential applications

It is still possible to use the developed control systems in this thesis manuscript for real-time optimal control of other flexible grid components, even though they have only been studied to control electric vehicles that are present on the distribution side. The electrification of the heating industry has resulted in a significant increase in the adoption of heat pumps, another new grid component. Sales in the EU reached a record-high 2.2 million units in 2021, a 34% increase from the previous year [84]. In addition, the REPowerEU target calls for installing 10 million hydronic heat pumps over the following five years [84]. Therefore, heat pumps can be envisioned as a key component of future smart grids. An optimization problem with heat pumps as the controlled elements can be modeled, similar to the optimization problems involving electric vehicles investigated in this thesis manuscript. The goal can be to reduce operating expenses over any chosen time horizon while ensuring stable grid operations and the satisfaction of heat pump owners. However, the objectives and goals pertaining to the operation of heat pumps must be translated into criticality values to align them with the adaptive multi-agent systems framework. This adjustment can be feasibly accomplished within a short-term time frame.

In the long-term, attention can be directed towards exploring the application of multi-armed bandit and adaptive multi-agent system theory in the domain of electrical transmission. The *optimal transmission switching problem* (OTSP) involves modifying the topology of a power grid in order to improve performance by managing the switching status of transmission lines. Increased computational requirements are challenges for this optimization problem as well [115]. The philosophy of adaptive multi-agent combined with machine learning may be applied to solve this problem in a decentralized manner and thus better prepare the electrical transmission systems to integrate increasing shares of grid energy storage systems and grid-connected distributed energy resources. The problem of *security-constrained optimal power flow* (SCOPF) for large-scale systems may also be tackled through decentralization of the control based on the methodologies studied in this manuscript [130].

Improvements in the system's functionality

It is possible to manage bus voltages through reactive power control. However, in this study, only active power control has been considered. The reasons are the fairly relatively resistive nature of distribution networks, the high cost of volt-ampere reactive (VAR) compensation mechanisms, and the fact that this is a preliminary study using combinatorial multi-armed bandits for optimal smart grid control with a focus on decentralization and operations performance under stochasticity. In the future, reactive power control can be integrated into the developed smart grid control system, especially for its application in the high voltage transmission side where reactive power control can be viewed as a relatively more viable option. Furthermore, frequency control functionality may also be implemented to provide frequency ancillary services to the electrical grid through vehicle-to-grid (V2G) technology. However, it must be noted that in the proposed system the cost of battery degradation due to continuous charging/discharging has not been penalized. The inclusion of this new penalization term in smart grid optimization problems becomes even more crucial when V2G technology is considered in the system. These functional modifications can be implemented within a relatively short to medium-term time frame.

Constraint optimization during exploration

During the exploration phase of reinforcement learning, the agent takes random or exploratory actions to learn about the environment and maximize its rewards. However, such exploratory behavior may result in the violation of constraints. This is an undesirable outcome for any system designed to control a smart grid in real-life as violating constraints can have a detrimental impact on the grid itself. The unpredictability introduced during the initial training phase of online reinforcement learning algorithms poses challenges for their practical implementation in real-life scenarios. Thus, it is crucial to also focus on the practical implementation of the proposed control system to ensure that the stable operation of the grid remains ensured even during the initial training phase of the learning agents.

To address the issue of constraint violations during exploration, novel methodologies such as *safe reinforcement learning* can be explored. In safe reinforcement learning, the satisfaction of the desired set of constraints is tried to be ensured even

during the exploration phase of the agents. This approach provides a means to mitigate the risk of violating critical constraints and maintains the stability and reliability of the smart grid. However, the application of such algorithms to smart grid optimization problems may not be straightforward. Given the complexity of this task, it can be considered a long-term perspective.

Asynchronicity in communication

The communication among all agents is assumed to be synchronous during the simulation-based experiments conducted in this manuscript. However, in reality, communication among different agents can be asynchronous, introducing potential challenges. Asynchronicity can lead to communication delays, miscommunication, and reduced collaboration among agents within the proposed adaptive multi-agent system. Therefore, it is crucial to investigate the performance of the decentralized control system in the presence of communication asynchronicity. If a significant negative impact on the system's performance is observed, it becomes necessary to enhance each agent's functionality to make it adaptable and resilient in the face of communication asynchronicity. Achieving this objective would involve evaluating the performance of the existing system and implementing significant modifications to various agent functionalities, which could present a challenging task. Therefore, this undertaking falls under the category of long-term perspectives.

Detailed economic viability studies

If the studied smart grid optimization problem requires the integration of vehicle-to-grid (V2G) functionality within the proposed combinatorial multi-armed bandit-based adaptive multi-agent smart grid control system, it becomes crucial to consider the cost associated with battery degradation. This consideration stems from the fact that discharging a battery can expedite its degradation, thereby exerting a notable impact on the economic viability of the solution. Thus, incorporating the cost of battery degradation becomes imperative when evaluating the economic feasibility of the proposed system in such contexts. It would be imperative to conduct comprehensive economic viability studies considering various scenarios encompassing the penetration of diverse distributed energy resources, the rate of cost reduction for different technologies, and the rate of efficiency improvement, among other factors. Undertaking such in-depth studies is essential prior to the selection of a smart grid solution for practical implementation, as it necessitates a thorough comprehension of the various variables involved. These thorough economic viability analyses can be completed in the medium to long-term time frame.

Detailed system analysis

A detailed analysis of the proposed adaptive multi-agent combinatorial multi-armed bandit system has been discussed in this manuscript in terms of its optimality, constraint satisfaction, fairness, and scalability. However, a number of sensitivity analyses can further be performed. This includes studying the impact of changing the total

number of daily decision instants or the length of the BRP's imbalance settlement period in the studied optimization problem on the optimality, constraints' satisfaction, and fairness of the system. Furthermore, the relationship between the choice of values of different learning parameters and criticality scaling parameters present in the proposed system can also be studied in detail. These studies can be performed in a short to medium-term time frame.

To compare the performance of the proposed algorithm uncontrolled strategy and MILP optimization strategy have been considered in this thesis manuscript. However, a variety of new decentralized smart grid control algorithms are being proposed using novel technologies such as reinforcement learning, peer-to-peer (P2P) trading, blockchain, and alternating direction method of multipliers (ADMM). These decentralized control algorithms can also be added to the mix of strategies that are being compared with the proposed system. This can be done in a medium to long-term time frame. This detailed comparative analysis will be an integral part of a future project named the PEPR TASE "TASTING" project (national project funded by the French National Research Agency (ANR), coordinated by G2Elab, 2023-2027).

Appendix A

Electric vehicle's reward function

The instantaneous electricity price $c(i)$ at instant i has a linear relationship with the instantaneous reward of an electric vehicle learning agent in the proposed decentralized optimal control system with combinatorial multi-armed bandit learning, as given in Equations (4.30) and (5.3). However other types of relationships are also possible, such as quadratic, exponential, etc. In this appendix, the impact of the modeled relationship between instantaneous electricity cost and EV's instantaneous reward on the system's performance is discussed. Following three relationships are studied here:

- **Linear relationship:** The instantaneous electric vehicle reward (in case the environment reward/criticality is null) is equal to $1 - c(i)$ at instant i.
- **Quadratic relationship:** The instantaneous electric vehicle reward (in case the environment reward/criticality is null) is equal to $1 - c(i)^2$ at instant i.
- **Logistic relationship:** The instantaneous electric vehicle reward (in case the environment reward/criticality is null) is equal to $\frac{1}{1+e^{-40(c(i)-0.5)}}$ at instant i.

All of the studied relationships are plotted in Figure A.1. It is evident that electric vehicle agents in the proposed system will observe different reward values depending on the selected relationship. Simulation experiments have been performed using the settings described in Section 5.3 with Thompson Sampling as the learning strategy.

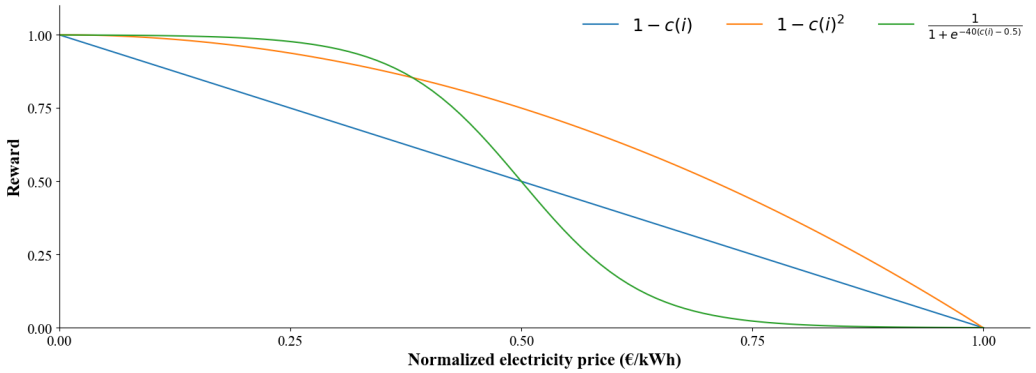

Figure A.1: Studied relationships between instantaneous electricity price and instantaneous electric vehicle's reward value.

The mean reward curve for each of the studied relationships during the training phase is shown in Figure A.2.

It can be observed in Figure A.2 that the system is converging to expected optimal solution in the same time irrespective of the modeled relationship. Furthermore, quadratic and logistic relationships are converging to a slightly higher average reward values only because they have a higher reward values at low electricity price instants compared to the linear relationship, as shown in Figure A.1. However, this does not mean that linear relationship is under-performing. The reward observed by an electric vehicle agent is only a "virtual" price that guides the agent towards an expected optimal policy. In practical life, the daily charging cost of each electric vehicle owner is calculated using the electricity price and not this "virtual" price. Here, the system converges to the same expected optimal solution for all of the three studied relationships. Thus, the system's performance remains unaffected both in terms of the convergence time and the quality of the solution. This is due to the fact that in each of the studied relationship, a distinction between the observed reward for each electricity price value can be made. This means the following general condition can be developed for any type of relationship between the electricity price and the instantaneous electric vehicle agent's reward to not have an impact on system's performance

No impact condition

$$\text{Reward for } c(x) > \text{Reward for } c(y) \quad \forall \quad c(x) < c(y) \tag{A.1}$$

If the above-stated condition is violated, i.e., the learning agent is not able to differentiate between reward values corresponding to different electricity prices (i.e., $c(x)$ and $c(y)$) then a degradation in the system's performance is expected. Thus, the satisfaction of Equation (A.1) guarantees that the system's performance remains near optimal for any type of modeled relationship between the instantaneous electric vehicle reward (in case the environment reward/criticality is null) and the instantaneous electricity price.

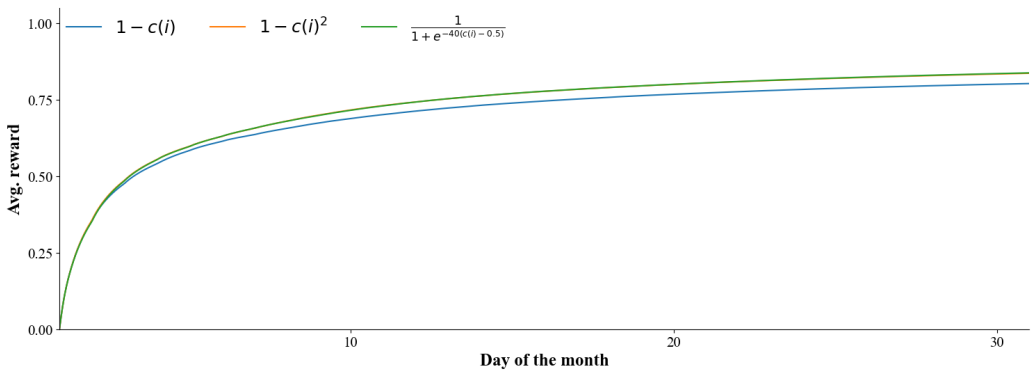

Figure A.2: Average learning reward of the system for the studied linear, quadratic, and logistic relationships.

Appendix B

Electric vehicles' charging policies

The daily charging policies picked by the three of the randomly selected decentralized electric vehicle charging agents in the CMAB-based learning system are shown in Figure B.1, Figure B.2, and Figure B.3.

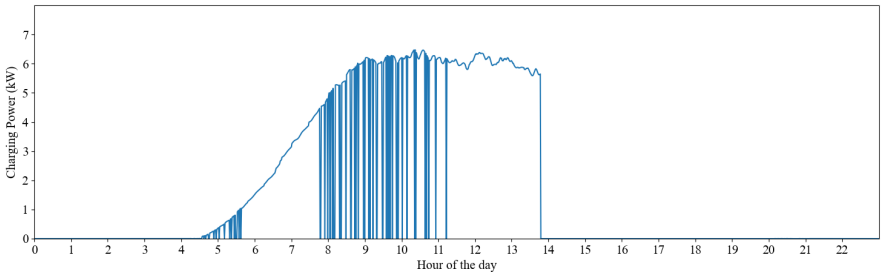

Figure B.1: Charging policy of the first randomly picked electric vehicle during one of the evaluation days.

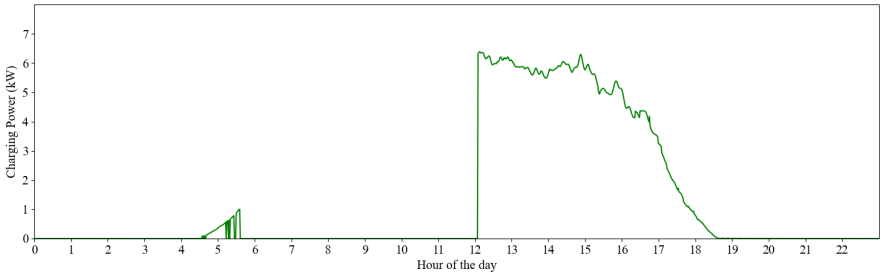

Figure B.2: Charging policy of the second randomly picked electric vehicle during one of the evaluation days.

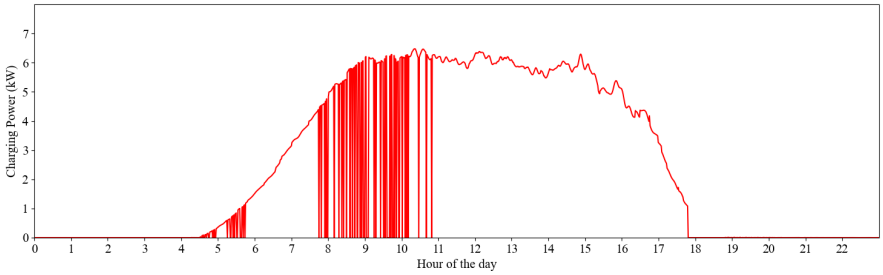

Figure B.3: Charging policy of the third randomly picked electric vehicle during one of the evaluation days.

Bibliography

[1] Agrawal, S., and Goyal, N. Analysis of thompson sampling for the multi-armed bandit problem. *CoRR abs/1111.1797* (2011).

[2] Akorede, M. F., Hizam, H., and Pouresmaeil, E. Distributed energy resources and benefits to the environment. *Renewable and Sustainable Energy Reviews 14*, 2 (2010), 724–734.

[3] Alanne, K., and Saari, A. Distributed energy generation and sustainable development. *Renewable and Sustainable Energy Reviews 10*, 6 (2006), 539–558.

[4] Alizadeh, F., and Goldfarb, D. Second-order cone programming. *Mathematical Programming 95* (12 2001).

[5] Allesiardo, R., Féraud, R., and Maillard, O.-A. The Non-stationary Stochastic Multi-armed Bandit Problem. *International Journal of Data Science and Analytics 3*, 4 (2017), 267–283.

[6] Andreu-Perez, J., Deligianni, F., Ravi, D., and Yang, G.-Z. Artificial intelligence and robotics, 2018.

[7] Applegate, D. L., Bixby, R. E., Chvátal, V., and Cook, W. J. *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, Princeton, 2007.

[8] Arnold, K., Gosling, J., and Holmes, D. *The Java programming language*. Addison Wesley Professional, 2005.

[9] ARRA. American recovery and reinvestment act (arra). https://www.energy.gov/recovery/arra, 2009. [Accessed: 15-Apr-2023].

[10] Asif, M., and Muneer, T. Energy supply, its demand and security issues for developed and emerging economies. *Renewable and Sustainable Energy Reviews 11*, 7 (2007), 1388–1413.

[11] Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning 47* (05 2002), 235–256.

[12] Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing 32*, 1 (2002), 48–77.

[13] Auer, P., and Ortner, R. Ucb revisited: Improved regret bounds for the stochastic multi-armed bandit problem. *Periodica Mathematica Hungarica 61* (09 2010), 55–65.

[14] Aygun, A. I., and Kamalasadan, S. Centralized charging approach to manage electric vehicle fleets for balanced grid. In *2022 IEEE International Conference on Power Electronics, Smart Grid, and Renewable Energy (PESGRE)* (2022), pp. 1–6.

[15] Bailey, D., and Wright, E. *Practical SCADA for Industry*. Newnes, Oxford, 2003.

[16] Baker, T., Gill, J., and Solovay, R. Relativizations of the $\mathcal{P} =? \mathcal{NP}$ question. *SIAM Journal on Computing 4*, 4 (1975), 431–442.

[17] Barlow, H. Unsupervised Learning. *Neural Computation 1*, 3 (09 1989), 295–311.

[18] Baroche, T. *Peer-to-peer electricity markets in power systems*. PhD thesis, Ecolé Normale Supérieure de Rennes , Jan. 2021.

[19] Becquerel, A. Memoire sur les effects d´electriques produits sous l´influence des rayons solaires. *Annalen der Physick und Chemie 54* (1841), 35–42.

[20] Belghache, E., Georgé, J.-P., and Gleizes, M.-P. Towards an adaptive multi-agent system for dynamic big data analytics. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress* (2016), pp. 753–758.

[21] Bernon, C., Gleizes, M.-P., Peyruqueou, S., and Picard, G. Adelfe: A methodology for adaptive multi-agent systems engineering. In *Engineering Societies in the Agents World III* (Berlin, Heidelberg, 2003), P. Petta, R. Tolksdorf, and F. Zambonelli, Eds., Springer Berlin Heidelberg, pp. 156–169.

[22] Berry, M. W., Mohamed, A., and Yap, B. W. *Supervised and Unsupervised Learning for Data Science*, 1st ed. Springer Publishing Company, Incorporated, 2019.

[23] Bertsimas, D., and Tsitsiklis, J. *Introduction to Linear Optimization*. Athena Scientific, 01 1998.

[24] Besson, L., and Kaufmann, E. Multi-player bandits revisited. *Algorithmic Learning Theory* (2017).

[25] Bhatti, A. R., Salam, Z., Aziz, M. J. B. A., Yee, K. P., and Ashique, R. H. Electric vehicles charging using photovoltaic: Status and technological review. *Renewable and Sustainable Energy Reviews 54* (2016), 34–47.

[26] Blanc-Rouchosse, J.-B., Blavette, A., Ben Ahmed, H., Camilleri, G., and Gleizes, M.-P. Multi-agent system for smart-grid control with commitment mismatch and congestion. In *2019 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)* (09 2019).

[27] BloombergNEF. Electric vehicle outlook 2021. https://about.bnef.com/electric-vehicle-outlook/, 2021. Accessed on April 15, 2023.

[28] Bonnefoi, R., Besson, L., Moy, C., Kaufmann, E., and Palicot, J. Multi-armed bandit learning in iot networks: Learning helps even in non-stationary settings. *12th EAI International Conference on Cognitive Radio Oriented Wireless Networks* (2018).

[29] Bonneville Power Administration . Total solar generation and total solar forecast in the BPA Balancing Authority (BPA) area. https://transmission.bpa.gov/Business/Operations/Wind/default.aspx, 2023. [Accessed: April 15, 2023].

[30] Borghetti, A., D'Ambrosio, C., Lodi, A., and Martello, S. An milp approach for short-term hydro scheduling and unit commitment with head-dependent reservoir. *IEEE Transactions on Power Systems 23*, 3 (2008), 1115–1124.

[31] Bouneffouf, D., and Féraud, R. Multi-armed bandit problem with known trend. *Neurocomputing 205* (2016), 16–21.

[32] Boutros, F., Doumiati, M., Olivier, J.-C., Mougharbel, I., and Kanaan, H. Y. Optimisation multi-objective d'un micro-réseau basée sur une nouvelle approche de modélisation avec un système de gestion d'énergie à temps-réel. In *Symposium Génie Electrique* (Lille, France, July 2023).

[33] Boutros, P. C., and Okey, A. B. Unsupervised pattern recognition: An introduction to the whys and wherefores of clustering microarray data. *Briefings in Bioinformatics 6*, 4 (12 2005), 331–343.

[34] Caruana, R., and Niculescu-Mizil, A. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning* (New York, NY, USA, 2006), ICML '06, Association for Computing Machinery, p. 161–168.

[35] Cerquides, J., Picard, G., and Rodríguez-Aguilar, J. A. Designing a marketplace for the trading and distribution of energy in the smart grid. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (2015), AAMAS '15, p. 1285–1293.

[36] Cesa-Bianchi, N., and Lugosi, G. Combinatorial bandits. *Journal of Computer and System Sciences 78*, 5 (2012), 1404–1422. JCSS Special Issue: Cloud Computing 2011.

[37] Chen, W., Wang, Y., and Yuan, Y. Combinatorial multi-armed bandit: General framework and applications. In *Proceedings of the 30th International Conference on Machine Learning* (Atlanta, Georgia, USA, 17–19 Jun 2013), S. Dasgupta and D. McAllester, Eds., vol. 28 of *Proceedings of Machine Learning Research*, PMLR, pp. 151–159.

[38] Chen, X., Nie, Y., and Li, N. Online residential demand response via contextual multi-armed bandits. *IEEE Control Systems Letters 5*, 2 (2021), 433–438.

[39] Chen, Z., Wu, L., and Fu, Y. Real-time price-based demand response management for residential appliances via stochastic optimization and robust optimization. *IEEE Transactions on Smart Grid 3*, 4 (2012), 1822–1831.

[40] Chérot, G., Le Goff Latimier, R., and Ben Ahmed, H. A real-time congestion control strategy in distribution networks. In *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)* (2021), pp. 1–5.

[41] CLEVER. Test-an-ev (test-en-elbil). http://mclabprojects.di.uniroma1.it/smarthgnew/Test-an-EV/?EV-code=, 2012. Accessed on April 15, 2023.

[42] Crama, Y. Combinatorial optimization models for production scheduling in automated manufacturing systems. *European Journal of Operational Research 99*, 1 (1997), 136–153.

[43] Crick, H. Q. P. The impact of climate change on birds. *Ibis 146*, s1 (2004), 48–56.

[44] Dakdouk, H., Tarazona, E., Alami, R., Feraud, R., Papadopoulos, G. Z., and Maillé, P. Reinforcement learning techniques for optimized channel hopping in ieee 802.15.4-tsch networks. *Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (10 2018), 99–107.

[45] Dallinger, D., and Wietschel, M. Grid integration of intermittent renewable energy sources using price-responsive plug-in electric vehicles. *Renewable and Sustainable Energy Reviews 16*, 5 (2012), 3370–3382.

[46] Danzi, P., Angjelichinoski, M., Stefanović, □., and Popovski, P. Distributed proportional-fairness control in microgrids via blockchain smart contracts. In *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)* (2017), pp. 45–51.

[47] Department of Energy . Benefits of demand response in electricity markets and recommendations for achieving them. https://eta.lbl.gov/publications/benefits-demand-response-electricity, 2006. [Accessed: April 15, 2023].

[48] Dhankhad, S., Mohammed, E., and Far, B. Supervised machine learning algorithms for credit card fraudulent transaction detection: A comparative study. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)* (2018), pp. 122–125.

[49] Di Marzo Serugendo, G., Gleizes, M.-P., and Karageorgos, A. Self-organization in multi-agent systems. *The Knowledge Engineering Review 20*, 2 (2005), 165–189.

[50] Di Silvestre, M. L., Favuzza, S., Riva Sanseverino, E., and Zizzo, G. How decarbonization, digitalization and decentralization are changing key power infrastructures. *Renewable and Sustainable Energy Reviews 93* (2018), 483–498.

[51] Diamond, S., and Boyd, S. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research 17*, 83 (2016), 1–5.

[52] DIgSILENT GmbH. PowerFactory. https://www.digsilent.de/products/powerfactory/, 2021. [Accessed: April 15, 2023].

[53] Dong, A., Le Goff Latimier, R., and Ben Ahmed, H. Asynchronous implementation of a region-based distributed optimal power flow algorithm. In *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)* (2022), pp. 1–5.

[54] Dorri, A., Kanhere, S. S., and Jurdak, R. Multi-agent systems: A survey. *IEEE Access 6* (2018), 28573–28593.

[55] Ecker, M., Nieto, N., Käbitz, S., Schmalstieg, J., Blanke, H., Warnecke, A., and Sauer, D. U. Calendar and cycle life study of li(nimnco)o2-based 18650 lithium-ion batteries. *Journal of Power Sources 248* (2014), 839–851.

[56] Ediger, V. S. . An integrated review and analysis of multi-energy transition from fossil fuels to renewables. *Energy Procedia 156* (2019), 2–6. 5th International Conference on Power and Energy Systems Engineering (CPESE 2018).

[57] Egbue, O., and Uko, C. Multi-agent approach to modeling and simulation of microgrid operation with vehicle-to-grid system. *The Electricity Journal 33*, 3 (2020), 106714.

[58] Electricity Consumers Resource Council (ELCON). The Economic Impacts of the August 2003 Blackout. https://elcon.org/wp-content/uploads/Economic20Impacts20of20August20200320Blackout1.pdf, 2004. [Accessed: April 15, 2023].

[59] European Commission. Smart grid mandate for europe: Task 1: Inventory of smart grid projects, task 2: Smart grid conceptual model, task 3: Roadmap for deployment of smart grid systems. https://ec.europa.eu/energy/sites/ener/files/documents/20110609_smartgrids_mandate_en.pdf, 2011. [Accessed: 15-Apr-2023].

[60] European Commission. Climate neutrality: Commission proposes transformation of eu economy and society to meet climate ambitions. https://ec.europa.eu/commission/presscorner/detail/en/IP_20_420, 2020. Accessed on April 15, 2023.

[61] European Parliament . Smart Grids: from innovation to deployment. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52011DC0202, 2011. [Accessed: April 15, 2023].

[62] European Parliament . EU Solar Energy Strategy. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A221%3AFIN&qid=1653034500503, 2022. [Accessed: April 15, 2023].

[63] European Parliament . Progress on competitiveness of clean energy technologies. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022DC0643&qid=1669913060946, 2022. [Accessed: April 15, 2023].

[64] European Union. Energy efficiency directive. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012L0027, 2012. Accessed on April 15, 2023.

[65] Fang, X., Misra, S., Xue, G., and Yang, D. Smart grid — the new and improved power grid: A survey. *IEEE Communications Surveys & Tutorials 14*, 4 (2012), 944–980.

[66] Fedjaev, J., Amamra, S.-A., and Francois, B. Linear programming based optimization tool for day ahead energy management of a lithium-ion battery for an industrial microgrid. *2016 IEEE International Power Electronics and Motion Control Conference (PEMC)* (2016), 406–411.

[67] Frank, S., and Rebennack, S. An introduction to optimal power flow: Theory, formulation, and examples. *IIE Transactions 48*, 12 (2016), 1172–1197.

[68] Fu, J., and Fu, Y. An adaptive multi-agent system for cost collaborative management in supply chains. *Engineering Applications of Artificial Intelligence 44* (2015), 91–100.

[69] Gai, Y., Krishnamachari, B., and Jain, R. Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation. In *2010 IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)* (2010), pp. 1–9.

[70] Garivier, A., and Moulines, E. On upper-confidence bound policies for switching bandit problems. In *Algorithmic Learning Theory* (2011), pp. 174–188.

[71] Gerard, H., Rivero Puente, E. I., and Six, D. Coordination between transmission and distribution system operators in the electricity sector: A conceptual framework. *Utilities Policy 50* (2018), 40–48.

[72] Goldberg, J. L. *Matrix theory with applications*. Springer New York, NY, 1991.

[73] Gowrisankaran, G., Reynolds, S. S., and Samano, M. Intermittency and the value of renewable energy. *Journal of Political Economy 124*, 4 (2016), 1187–1234.

[74] Gray, M., and Morsi, W. On the impact of single-phase plug-in electric vehicles charging and rooftop solar photovoltaic on distribution transformer aging. *Electric Power Systems Research 148* (2017), 202–209.

[75] Habibidoost, M., and Bathaee, S. M. T. A self-supporting approach to ev agent participation in smart grid. *International Journal of Electrical Power & Energy Systems 99* (2018), 394–403.

[76] Hadjipaschalis, I., Poullikkas, A., and Efthimiou, V. Overview of current and future energy storage technologies for electric power applications. *Renewable and Sustainable Energy Reviews 13*, 6 (2009), 1513–1522.

[77] Hartmanis, J., and Stearns, R. E. On the computational complexity of algorithms. *Transactions of the American Mathematical Society 117* (1965), 285–306.

[78] Hledik, R. How green is the smart grid? *The Electricity Journal 22*, 3 (2009), 29–41.

[79] Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association 58*, 301 (1963), 13–30.

[80] Holweger, J., Pena-Bello, A., Jeannin, N., Ballif, C., and Wyrsch, N. Distributed flexibility as a cost-effective alternative to grid reinforcement. *Sustainable Energy, Grids and Networks* (2023), 101041.

[81] Hu, J., Morais, H., Lind, M., and Bindner, H. W. Multi-agent based modeling for electric vehicle integration in a distribution network operation. *Electric Power Systems Research 136* (2016), 341–351.

[82] Huang, H., Cai, Y., Xu, H., and Yu, H. A multiagent minority-game-based demand-response management of smart buildings toward peak load reduction. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 36*, 4 (2017), 573–585.

[83] Ikegami, T., Ogimoto, K., Yano, H., Kudo, K., and Iguchi, H. Balancing power supply-demand by controlled charging of numerous electric vehicles. In *2012 IEEE International Electric Vehicle Conference* (2012), pp. 1–8.

[84] International Energy Agency (IEA). Is the European Union on track to meet its REPowerEU goals? https://www.iea.org/reports/is-the-european-union-on-track-to-meet-its-repowereu-goals, 2022. [Accessed: April 15, 2023].

[85] International Energy Agency. Technology roadmap smart grid. https://www.iea.org/reports/technology-roadmap-smart-grids, 2011. [Accessed: April 15, 2023].

[86] International Energy Agency. Global energy & co2 status report 2019. https://www.iea.org/reports/global-energy-co2-status-report-2019, 2019. Accessed on April 15, 2023.

[87] International Energy Agency. Global energy review 2021. https://www.iea.org/reports/global-energy-review-2021/co2-emissions, 2021. Accessed on April 15, 2023.

[88] International Energy Agency. Global ev outlook 2021. https://www.iea.org/reports/global-ev-outlook-2021, 2021. [Accessed: 15-Apr-2023].

[89] International Energy Agency. Renewables 2021: Analysis and forecast to 2026. https://www.iea.org/reports/renewables-2021, 2021. Accessed on April 15, 2023.

[90] International Energy Agency. Smart Grids. https://www.iea.org/reports/smart-grids, 2022. [Accessed: April 15, 2023].

[91] International Monetary Fund. Global financial stability report: Lower for longer. https://www.imf.org/en/Publications/GFSR/Issues/2019/10/01/global-financial-stability-report-october-2019, 2019. Accessed on April 15, 2023.

[92] International Renewable Energy Agency . Renewable capacity statistics. https://www.irena.org/Publications/2023/Mar/Renewable-capacity-statistics-2023, 2023. [Accessed: April 15, 2023].

[93] International Renewable Energy Agency. Global renewables outlook: Energy transformation 2050. https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2020/Apr/IRENA_Global_Renewables_Outlook_2020.pdf, 2020. [Accessed: 15-Apr-2023].

[94] Jain, S., Narayanaswamy, B., and Narahari, Y. A multiarmed bandit incentive mechanism for crowdsourcing demand response in smart grids. *Proceedings of the AAAI Conference on Artificial Intelligence 28* (Jun. 2014).

[95] Jones, C. B., Lave, M., Vining, W., and Garcia, B. M. Uncontrolled electric vehicle charging impacts on distribution electric power systems with primarily residential, commercial or industrial loads. *Energies 14*, 6 (2021).

[96] Joskow, P. L., and Wolfram, C. D. Dynamic pricing of electricity. *American Economic Review 102*, 3 (May 2012), 381–85.

[97] Joyce, J. Bayes' Theorem. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., Fall 2021 ed. Metaphysics Research Lab, Stanford University, 2021.

[98] Kane, M. France: More than 315,000 plug-in electric cars were sold in 2021. https://insideevs.com/news/561101/france-plugin-car-sales-2021/, 2022. Accessed: 2023-02-22.

[99] Kara, E. C., Macdonald, J. S., Black, D., Bérges, M., Hug, G., and Kiliccote, S. Estimating the benefits of electric vehicle smart charging at non-residential locations: A data-driven approach. *Applied Energy 155* (2015), 515–525.

[100] Karfopoulos, E. L., and Hatziargyriou, N. D. A multi-agent system for controlled charging of a large population of electric vehicles. *IEEE Transactions on Power Systems 28*, 2 (2013), 1196–1204.

[101] Katiraei, F., and Agüero, J. R. Solar pv integration challenges. *IEEE Power and Energy Magazine 9*, 3 (2011), 62–71.

[102] Kim, J.-H., Shim, H.-S., Kim, H.-S., Jung, M.-J., Choi, I.-H., and Kim, J.-O. A cooperative multi-agent system and its real time application to robot soccer. In *Proceedings of International Conference on Robotics and Automation* (1997), vol. 1, pp. 638–643 vol.1.

[103] Kintner-Meyer, M., Schneider, K., and Pratt, R. Impacts assessment of plug-in hybrid vehicles on electric utilities and regional us power grids: Part 1: Technical analysis. *Online Journal of EUEC* (01 2007).

[104] Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Sallab, A. A. A., Yogamani, S., and Pérez, P. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems 23*, 6 (2022), 4909–4926.

[105] Koliou, E., Eid, C., Chaves-Ávila, J. P., and Hakvoort, R. A. Demand response in liberalized electricity markets: Analysis of aggregated load participation in the german balancing mechanism. *Energy 71* (2014), 245–254.

[106] Kumar, R., Sharma, D., and Sadu, A. A hybrid multi-agent based particle swarm optimization algorithm for economic power dispatch. *International Journal of Electrical Power & Energy Systems 33*, 1 (2011), 115–123.

[107] Kundur, P., and Malik, O. *Power System Stability and Control, Second Edition*. The EPRI power system engineering series. McGraw-Hill Education, 2022.

[108] Kurevska, L., Sauhats, A., Junghans, G., and Lavrinovcs, V. Harmonization of imbalance settlement period across europe: the curious case of baltic energy markets. In *2019 IEEE 60th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON)* (2019), pp. 1–5.

[109] Lai, T., and Robbins, H. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics 6*, 1 (1985), 4–22.

[110] Lauri, F., Basso, G., Zhu, J., Roche, R., Hilaire, V., and Koukam, A. Managing power flows in microgrids using multi-agent reinforcement learning. *Agent Technologies for Energy Systems* (05 2013).

[111] Le Goff Latimier, R. *Gestion et dimensionnement d'une flotte de véhicules électriques associée à une centrale photovoltaïque : co-optimisation stochastique et distribuée*. PhD thesis, Université Paris Saclay (COmUE), Sept. 2016.

[112] Le Goff Latimier, R., Chérot, G., and Ben Ahmed, H. Online learning for distributed optimal control of an electric vehicle fleet. *Electric Power Systems Research 212* (2022), 108330.

[113] Leo, R., Milton, R., and Morais, A. A. Autonomous energy management of a micro-grid using multi agent system. *Indian Journal of Science and Technology 9* (04 2016).

[114] Li, L., Chu, W., Langford, J., and Schapire, R. E. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web* (New York, NY, USA, 2010), WWW '10, Association for Computing Machinery, p. 661–670.

[115] Li, X., Balasubramanian, P., Sahraei-Ardakani, M., Abdi-Khorsand, M., Hedman, K., and Podmore, R. Real-time contingency analysis with corrective transmission switching - part i: Methodology. *IEEE Power & Energy Society General Meeting* (04 2016).

[116] Liu, C., Chai, K. K., Zhang, X., Lau, E. T., and Chen, Y. Adaptive blockchain-based electric vehicle participation scheme in smart grid platform. *IEEE Access 6* (2018), 25657–25665.

[117] Liu, Y., Zhou, P., Yang, L., Wu, Y., Xu, Z., Liu, K., and Wang, X. Privacy-preserving context-based electric vehicle dispatching for energy scheduling in microgrids: An online learning approach. *IEEE Transactions on Emerging Topics in Computational Intelligence 6*, 3 (2022), 462–478.

[118] Lo, C.-H., and Ansari, N. Decentralized controls and communications for autonomous distribution networks in smart grid. *IEEE Transactions on Smart Grid 4*, 1 (2013), 66–77.

[119] Lopes, J. A. P., Soares, F. J., and Almeida, P. M. R. Integration of electric vehicles in the electric power system. *Proceedings of the IEEE 99*, 1 (2011), 168–183.

[120] Lu, T., Pal, D., and Pal, M. Contextual multi-armed bandits. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010), Y. W. Teh and M. Titterington, Eds., vol. 9 of *Proceedings of Machine Learning Research*, PMLR, pp. 485–492.

[121] Maghsudi, S., and Hossain, E. Multi-armed bandits with application to 5g small cells. *IEEE Wireless Communications 23*, 3 (2016), 64–73.

[122] Magnanti, T. L. Combinatorial optimization and vehicle fleet planning: Perspectives and prospects. *Networks 11*, 2 (1981).

[123] Mahale, R. A., and Chavan, P. S. D. A survey : Evolutionary and swarm based bio-inspired optimization algorithms. *International Journal of Scientific and Research Publications 2* (2012).

[124] Massoud Amin, S., and Wollenberg, B. Toward a smart grid: power delivery for the 21st century. *IEEE Power and Energy Magazine 3*, 5 (2005), 34–41.

[125] Mets, K., De Turck, F., and Develder, C. Distributed smart charging of electric vehicles for balancing wind energy. In *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm)* (2012), pp. 133–138.

[126] Mocci, S., Natale, N., Pilo, F., and Ruggeri, S. Multi-agent control system to coordinate optimal electric vehicles charging and demand response actions in active distribution networks. In *3rd Renewable Power Generation Conference (RPG 2014)* (2014), pp. 1–6.

[127] Mohamed, A., Salehi, V., Ma, T., and Mohammed, O. Real-time energy management algorithm for plug-in hybrid electric vehicle charging parks involving sustainable energy. *IEEE Transactions on Sustainable Energy 5*, 2 (2014), 577–586.

[128] Molena, F. Remarkable Weather of 1911. http://www.logboekweer.nl/Actueel/PopularMechanicsMagazine1912.pdf, 1911. [Accessed: April 15, 2023].

[129] Mollah, M. B., Zhao, J., Niyato, D., Lam, K.-Y., Zhang, X., Ghias, A. M. Y. M., Koh, L. H., and Yang, L. Blockchain for future smart grid: A comprehensive survey. *IEEE Internet of Things Journal 8*, 1 (2021), 18–43.

[130] Monticelli, A., Pereira, M. V. F., and Granville, S. Security-constrained optimal power flow with post-contingency corrective rescheduling. *IEEE Transactions on Power Systems 2*, 1 (1987), 175–180.

[131] Munsing, E., Mather, J., and Moura, S. Blockchains for decentralized optimization of energy resources in microgrid networks. In *2017 IEEE Conference on Control Technology and Applications (CCTA)* (08 2017), pp. 2164–2171.

[132] Nasteski, V. An overview of the supervised machine learning methods. *HORIZONS.B 4* (12 2017), 51–62.

[133] National Renewable Energy Laboratory. NREL Solar Data. https://www.nrel.gov/grid/solar-resource/solar-data.html, 2021. [Accessed: April 15, 2023].

[134] Neyman, J. Frequentist probability and frequentist statistics. *Synthese 36*, 1 (1977), 97–131.

[135] Nguyen, T.-L., Tran, Q.-T., Caire, R., Wang, Y., Besanger, Y., and Luu, N.-A. Distributed optimal power flow and the multi-agent system for the realization in cyber-physical system. *Electric Power Systems Research 192* (2021), 107007.

[136] Nguyen, T. N., and Müsgens, F. What drives the accuracy of pv output forecasts? *Applied Energy 323* (2022), 119603.

[137] Nunna, H. S. V. S. K., Battula, S., Doolla, S., and Srinivasan, D. Energy management in smart distribution systems with vehicle-to-grid integrated microgrids. *IEEE Transactions on Smart Grid 9*, 5 (2018), 4004–4016.

[138] Nwana, H. S. Software agents: an overview. *The Knowledge Engineering Review 11* (1996), 205 – 244.

[139] Official Journal of the European Communities. Council Resoulution of 25 July 1983 on framework programmes for Community research, development and demonstration activities and a first framework programme 1984 to 1987. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31983Y0804(01), 1983. [Accessed: April 15, 2023].

[140] Ontañón, S. Combinatorial multi-armed bandits for real-time strategy games. *Journal of Artificial Intelligence Research (JAIR)* (2017).

[141] Papadimitriou, C. H., and Steiglitz, K. *Combinatorial Optimization : Algorithms and Complexity*. Dover Publications, July 1998.

[142] Papadopoulos, P., Jenkins, N., Cipcigan, L. M., Grau, I., and Zabala, E. Coordination of the charging of electric vehicles using a multi-agent system. *IEEE Transactions on Smart Grid 4*, 4 (2013), 1802–1809.

[143] Perles, A., Crasnier, F., and Georgé, J.-P. Amak - a framework for developing robust and open adaptive multi-agent systems. In *Highlights of Practical Applications of Agents, Multi-Agent Systems, and Complexity: The PAAMS Collection* (Cham, 2018), J. Bajo, J. M. Corchado, E. M. Navarro Martínez, E. Osaba Icedo, P. Mathieu, P. Hoffa-Dąbrowska, E. del Val, S. Giroux, A. J. Castro, N. Sánchez-Pi, V. Julián, R. A. Silveira, A. Fernández, R. Unland, and R. Fuentes-Fernández, Eds., Springer International Publishing, pp. 468–479.

[144] Petit, M., and Hennebel, M. Ev smart charging in collective residential buildings: the bienvenu project. In *2019 IEEE Milan PowerTech* (2019), pp. 1–6.

[145] Petrusev, A., Putratama, M. A., Rigo-Mariani, R., Debusschere, V., Reignier, P., and Hadjsaid, N. Reinforcement learning for robust voltage control in distribution grids under uncertainties. *Sustainable Energy, Grids and Networks 33* (2023), 100959.

[146] Phuangpornpitak, N., and Tia, S. Opportunities and challenges of integrating renewable energy in smart grid system. *Energy Procedia 34* (2013), 282–290. 10th Eco-Energy and Materials Science and Engineering Symposium.

[147] Pipattanasomporn, M., Feroze, H., and Rahman, S. Multi-agent systems in a distributed smart grid: Design and implementation. In *2009 IEEE/PES Power Systems Conference and Exposition* (2009), pp. 1–8.

[148] Prencipe, A., Davies, A., and Hobday, M. *The Business of Systems Integration*. Oxford University Press, 11 2003.

[149] Quddus, M. A., Shahvari, O., Marufuzzaman, M., Usher, J. M., and Jaradat, R. A collaborative energy sharing optimization model among electric vehicle charging stations, commercial buildings, and power grid. *Applied Energy 229* (2018), 841–857.

[150] Radhakrishnan, B. M., Srinivasan, D., and Mehta, R. Fuzzy-based multi-agent system for distributed energy management in smart grids. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 24*, 05 (2016), 781–803.

[151] Ray, S. A quick review of machine learning algorithms. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (2019), pp. 35–39.

[152] Renuka, D. K., Hamsapriya, T., Chakkaravarthi, M. R., and Surya, P. L. Spam classification based on supervised learning using machine learning techniques. In *2011 International Conference on Process Automation, Control and Computing* (2011), pp. 1–7.

[153] Rivero, E., Sebastian-Viana, M., Ulian, A., and Stromsather, J. The evolvDSO project: Key services for the evolution of DSOs roles. *23rd International Conference on Electricity Distri* (Sept. 2015).

[154] Roberts, B. P., and Sandberg, C. The role of energy storage in development of smart grids. *Proceedings of the IEEE 99*, 6 (2011), 1139–1144.

[155] Roozbehani, M., Dahleh, M., and Mitter, S. Dynamic pricing and stabilization of supply and demand in modern electric power grids. In *2010 First IEEE International Conference on Smart Grid Communications* (2010), pp. 543–548.

[156] Roszczypala, D., Batard, C., Ginot, N., and Poitiers, F. Optimisation de charge de véhicules électriques au sein d'un micro-réseau avec production d'énergie renouvelable, et stockage fixe. In *Symposium de Génie Electrique* (Nancy, France, July 2018), Université de Lorraine [UL].

[157] Sahar, F. Machine-learning techniques for customer retention: A comparative study. *International Journal of Advanced Computer Science and Applications 9* (01 2018).

[158] Salkin, H. M., and De Kluyver, C. A. The knapsack problem: A survey. *Naval Research Logistics Quarterly 22*, 1 (1975), 127–144.

[159] Schneider, K. P., Mather, B. A., Pal, B. C., Ten, C. W., Shirek, G. J., Zhu, H., Fuller, J. C., Pereira, J. L. R., Ochoa, L. F., de Araujo, L. R., Dugan, R. C., Matthias, S., Paudyal, S., McDermott, T. E., and Kersting, W. Analytic considerations and design basis for the ieee distribution test feeders. *IEEE Transactions on Power Systems PP*, 99 (2017), 1–1.

[160] Shirazi, E., and Jadid, S. A multiagent design for self-healing in electric power distribution systems. *Electric Power Systems Research 171* (02 2019).

[161] Slivkins, A. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning 12*, 1-2 (2019), 1–286.

[162] Smil, V. Energy in the twentieth century: Resources, conversions, costs, uses, and consequences. *Annual Review of Energy and the Environment 25*, 1 (2000), 21–51.

[163] Sobri, S., Koohi-Kamali, S., and Rahim, N. A. Solar photovoltaic generation forecasting methods: A review. *Energy Conversion and Management 156* (2018), 459–497.

[164] Strens, M. J. A. A bayesian framework for reinforcement learning. In *Proceedings of the Seventeenth International Conference on Machine Learning* (San Francisco, CA, USA, 2000), ICML '00, Morgan Kaufmann Publishers Inc., p. 943–950.

[165] Sueyoshi, T., and Goto, M. Photovoltaic power stations in germany and the united states: A comparative study by data envelopment analysis. *Energy Economics 42* (2014), 271–288.

[166] Sutton, R. S., and Barto, A. G. *Reinforcement learning: An introduction*. MIT press, 2018.

[167] Swaminathan, B., Debusschere, V., and Caire, R. Short-term active distribution network operation with convex formulations of power flow. In *2017 IEEE Manchester PowerTech* (2017), pp. 1–6.

[168] Taes, S. Electric cars in history. https://www.europeana.eu/en/blog/electric-cars-in-history, 2021. [Accessed: April 15, 2023].

[169] Tavakoli, A., Saha, S., Arif, M. T., Haque, M. E., Mendis, N., and Oo, A. M. Impacts of grid integration of solar pv and electric vehicle on grid stability, power quality and energy economics: a review. *IET Energy Systems Integration 2*, 3 (2020), 243–260.

[170] Teodorovic´, D. Transport modeling by multi-agent systems: a swarm intelligence approach. *Transportation Planning and Technology 26*, 4 (2003), 289–312.

[171] Thompson, W. R. ON THE LIKELIHOOD THAT ONE UNKNOWN PROBABILITY EXCEEDS ANOTHER IN VIEW OF THE EVIDENCE OF TWO SAMPLES. *Biometrika 25*, 3-4 (12 1933), 285–294.

[172] Thurner, L., Scheidler, A., Schafer, F., Menke, J. H., Dollichon, J., Meier, F., Meinecke, S., and Braun, M. pandapower - an open source python tool for convenient modeling, analysis and optimization of electric power systems. *IEEE Transactions on Power Systems* (2018).

[173] Toit, J. D., Davimes, R., Mohamed, A. J., Patel, K., and Nye, J. M. Customer segmentation using unsupervised learning on daily energy load profiles. *Journal of Advances in Information Technology 7* (2016), 69–75.

[174] Tol, R. S. J. The economic impacts of climate change. *Review of Environmental Economics and Policy 12*, 1 (2018), 4–25.

[175] Ucer, E., Kisacikoglu, M. C., Yuksel, M., and Gurbuz, A. C. An internet-inspired proportional fair ev charging control method. *IEEE Systems Journal 13*, 4 (2019), 4292–4302.

[176] United States Environmental Protection Agency (USEPA). Centralized Generation of Electricity and its Impacts on the Environment. https://www.epa.gov/energy/centralized-generation-electricity-and-its-impacts-environment, 2023. [Accessed: April 15, 2023].

[177] U.S. Department of Energy. Smart grid policy statement. "https://www.smartgrid.gov/files/SG_Implementation_Guidance_01-20-11_Final_0.pdf", 2010. [Accessed: 15-Apr-2023].

[178] U.S. Federal Energy Regulatory Commission. Order no. 1000: Transmission planning and cost allocation by transmission owning and operating public utilities. https://www.ferc.gov/legal/maj-ord-reg/land-docs/order1000-final.pdf, 2011. [Accessed: 15-Apr-2023].

[179] van der Kam, M., and van Sark, W. Smart charging of electric vehicles with photovoltaic power and vehicle-to-grid technology in a microgrid; a case study. *Applied Energy 152* (2015), 20–30.

[180] Van der Kam, M., and Van Sark, W. Smart charging of electric vehicles with photovoltaic power and vehicle-to-grid technology in a microgrid; a case study. *Applied Energy 152* (2015), 20–30.

[181] Van Rossum, G., and Drake, F. L. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.

[182] Vandael, S., Boucké, N., Holvoet, T., De Craemer, K., and Deconinck, G. Decentralized coordination of plug-in hybrid vehicles for imbalance reduction in a smart grid. *International Foundation for Autonomous Agents and Multiagent Systems* (2011).

[183] Vermorel, J., and Mohri, M. Multi-armed bandit algorithms and empirical evaluation. In *Machine Learning: ECML 2005* (Berlin, Heidelberg, 2005), J. Gama, R. Camacho, P. B. Brazdil, A. M. Jorge, and L. Torgo, Eds., Springer Berlin Heidelberg, pp. 437–448.

[184] Viana, A., and Pedroso, J. P. A new milp-based approach for unit commitment in power production planning. *International Journal of Electrical Power & Energy Systems 44*, 1 (2013), 997–1005.

[185] Vinyals, O., Ewalds, T., Bartunov, S., Georgiev, P., Vezhnevets, A. S., Yeo, M., Makhzani, A., Küttler, H., Agapiou, J., Schrittwieser, J., Quan, J., Gaffney, S., Petersen, S., Simonyan, K., Schaul, T., van Hasselt, H., Silver, D., Lillicrap, T., Calderone, K., Keet, P., Brunasso, A., Lawrence, D., Ekermo, A., Repp, J., and Tsing, R. Starcraft ii: A new challenge for reinforcement learning. *ArXiv* (2017).

[186] Wang, K., Gu, L., He, X., Guo, S., Sun, Y., Vinel, A., and Shen, J. Distributed energy management for vehicle-to-grid networks. *IEEE Network 31*, 2 (2017), 22–28.

[187] Wang, L., and Chen, B. Distributed control for large-scale plug-in electric vehicle charging with a consensus algorithm. *International Journal of Electrical Power & Energy Systems 109* (2019), 369–383.

[188] Wang, S., and Chen, W. Thompson sampling for combinatorial semi-bandits. *Proceedings of the 35th International Conference on Machine Learning* (2018).

[189] Wang, Y., Nguyen, T.-L., Xu, Y., Tran, Q.-T., and Caire, R. Peer-to-peer control for networked microgrids: Multi-layer and multi-agent architecture design. *IEEE Transactions on Smart Grid 11*, 6 (2020), 4688–4699.

[190] Wang, Y., Yang, Z., Mourshed, M., Guo, Y., Niu, Q., and Zhu, X. Demand side management of plug-in electric vehicles and coordinated unit commitment: A novel parallel competitive swarm optimization method. *Energy Conversion and Management 196* (2019), 935–949.

[191] Weber, M., Welling, M., and Perona, P. Unsupervised learning of models for recognition. In *Computer Vision - ECCV 2000* (Berlin, Heidelberg, 2000), Springer Berlin Heidelberg, pp. 18–32.

[192] Wen, X., Abbes, D., and Francois, B. Day-ahead generation planning and power reserve allocation with a flexible storage strategy. In *CIRED 2020 Berlin Workshop (CIRED 2020)* (2020), vol. 2020, pp. 553–556.

[193] Wen, X., Abbes, D., and Francois, B. Stochastic optimization for security-constrained day-ahead operational planning under pv production uncertainties: Reduction analysis of operating economic costs and carbon emissions. *IEEE Access 9* (2021), 97039–97052.

[194] Wi, Y.-M., Lee, J.-U., and Joo, S.-K. Electric vehicle charging method for smart homes/buildings with a photovoltaic system. *IEEE Transactions on Consumer Electronics 59*, 2 (2013), 323–328.

[195] Wi, Y.-M., Lee, J.-U., and Joo, S.-K. Electric vehicle charging method for smart homes/buildings with a photovoltaic system. *IEEE Transactions on Consumer Electronics 59*, 2 (2013), 323–328.

[196] Yan, T., He, Z., Zhao, N., Zhang, Z., and Zhang, T. Coordinated charging and discharging of electric vehicles for power imbalance mitigation. In *2021 International Conference on Power System Technology (POWERCON)* (2021), pp. 778–783.

[197] Yang, B., Wang, L.-f., Liao, C.-l., and Ji, L. Coordinated charging method of electric vehicles to deal with uncertainty factors. In *2014 IEEE Conference and Expo Transportation Electrification Asia-Pacific (ITEC Asia-Pacific)* (2014), pp. 1–6.

[198] Yang, Q., and Wang, H. Blockchain-empowered socially optimal transactive energy system: Framework and implementation. *IEEE Transactions on Industrial Informatics 17*, 5 (2021), 3122–3132.

[199] Yu, Z., Xu, Y., and Tong, L. Large scale charging of electric vehicles: A multi-armed bandit approach. *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (2015), 389–395.

[200] Zafar, S., Blavette, A., Camilleri, G., Ben Ahmed, H., and Prince Agbodjan, J.-J. Decentralized optimal management of a large-scale ev fleet: Optimality and computational complexity comparison between an adaptive mas and milp. *International Journal of Electrical Power & Energy Systems 147* (2023), 108861.

[201] Zhang, Z., Zhang, D., and Qiu, R. C. Deep reinforcement learning for power system applications: An overview. *CSEE Journal of Power and Energy Systems 6*, 1 (2020), 213–225.

[202] Zhao, Q., Liu, Y., Yao, W., and Yao, Y. Hourly rainfall forecast model using supervised learning algorithm. *IEEE Transactions on Geoscience and Remote Sensing 60* (2022), 1–9.

[203] Zheng, Y., Hill, D. J., and Dong, Z. Y. Multi-agent optimal allocation of energy storage systems in distribution systems. *IEEE Transactions on Sustainable Energy 8*, 4 (2017), 1715–1725.

[204] Zhou, K., Cheng, L., Wen, L., Lu, X., and Ding, T. A coordinated charging scheduling method for electric vehicles considering different charging demands. *Energy 213* (2020), 118882.

[205] Zhu, J., Song, Y., Jiang, D., and Song, H. Multi-armed bandit channel access scheme with cognitive radio technology in wireless sensor networks for the internet of things. *IEEE Access 4* (2016), 4609–4617.

[206] Zhu, K., and Zhang, T. Deep reinforcement learning based mobile robot navigation: A review. *Tsinghua Science and Technology 26*, 5 (2021), 674–691.

[207] Zishan, A. A., Haji, M. M., and Ardakanian, O. Adaptive congestion control for electric vehicle charging in the smart grid. *IEEE Transactions on Smart Grid 12*, 3 (2021), 2439–2449.

COLLEGE MATHS, TELECOMS
DOCTORAL INFORMATIQUE, SIGNAL
BRETAGNE SYSTEMES, ELECTRONIQUE

école normale supérieure rennes

Titre : Gestion optimisée d'un réseau de distribution actif par AMAS couplé à la méthode RL des bandits

Mot clés : Contrôle décentralisé, AMAS, Bandit manchot, MARL, recharge intelligente des VE

Résumé : Les systèmes électriques modernes évoluent avec l'introduction des ressources énergétiques distribuées et des véhicules électriques, promettant la durabilité. Cependant, l'intégration non contrôlée de ces technologies dans les réseaux électriques existants peut entraîner des déséquilibres en temps réel et des problèmes de le pic de la demande. Le renforcement traditionnel du réseau présente des inconvénients, notamment des préoccupations liées au coût et au temps de déploiement. Des solutions flexibles, rendues possibles par la digitalisation du réseau, offrent une alternative en contrôlant dynamiquement les éléments du réseau. Cependant, l'optimisation de ces solutions pour les différents acteurs du marché est complexe, et les approches centralisées peuvent avoir du mal à gérer en temps réel de grands réseaux intelligents. Cette thèse aborde ces défis en développant un système décentralisé utilisant des systèmes multi-agents adaptatifs pour le contrôle en temps réel des entités flexibles dans les réseaux de distribution. Des expériences de simulation valident son efficacité pour surmonter les problèmes de centralisation. De plus, l'intégration de l'apprentissage combinatoire à bandit manchot améliore les performances dans des environnements stochastiques. Cette recherche propose une approche prometteuse pour l'optimisation de grands réseaux intelligents alors qu'ils s'adaptent aux évolutions du paysage énergétique.

Title: Optimized management of an active distribution network using AMAS combined with the RL bandit method

Keywords: Decentralized control, AMAS, Multi-armed bandits, MARL, EVs smart charging

Abstract: Modern electrical power systems are evolving with the introduction of distributed energy resources and electric vehicles, promising sustainability. However, the uncontrolled integration of these technologies into legacy power grids can lead to real-time imbalances and peak load issues. Traditional grid reinforcement has drawbacks, including cost and deployment time concerns. Flexible solutions, enabled by grid digitization, offer an alternative by dynamically controlling grid elements. Yet, optimizing these solutions for diverse market actors is complex, and centralized approaches may struggle to manage large-scale smart grids in real-time. This thesis addresses these challenges by developing a decentralized system using adaptive multi-agent systems for real-time control of flexible entities in distribution grids. Simulation experiments validate its effectiveness in overcoming centralization issues. Furthermore, integrating combinatorial multi-armed bandit learning enhances performance in stochastic environments. This research offers a promising approach to optimizing large-scale smart grids as they adapt to evolving energy landscapes.