UNIVERSITY OF TWENTE.

Faculty of Engineering Technology

master thesis

Sustainable Energy Technology

Applying game-theory to evaluate incentives for cooperative behaviour between households within the distribution grid

Geert Buis

Governance and Technology for Sustainability
25th of August 2023

Graduation committee
Frans Coenen
Athanasios Votsis
Gerwin Hoogsteen

Master Graduation Assignment for the master's programme Sustainable Energy Technology

University of Twente
Enschede, July 2023

Acknowledgements

My thanks goes out to the graduation committee, who showed great patience in helping me in finding a suitable research topic. Their enthusiasm for looking beyond their respective fields of study is what made this interdisciplinary work possible. I would like to thank Gerwin and Frans for their interest in my work and their feedback throughout the writing process. Special appreciation is reserved for Athanasios, who helped me see the real-life political relevance of my work, and encouraged me to trust my own judgements.

I would also like to thank Teresa, who was been an endless source of support, as well as my family, my friends, and my colleagues at AmperaPark. They spent much time listening to my worries and complaints, and their encouragement is what enabled me to take this work to the next level.

Abstract—**This thesis presents a novel perspective on addressing congestion within distribution grids for electricity. Departing from conventional approaches from literature, which often rely on complex market or control structures, this study applies common-pool resource theory to the distribution grid, treating it as a socio-technical system. By applying game theory and reinforcement learning, the study models interactions among self-interested players sharing grid capacity, yielding insights into congestion management strategies. The findings underscore the potential of community-based approaches, and emphasise the need for institutional facilitation through the provision of data and the redistribution of funds. The study advocates for a holistic approach to congestion management that aligns with sustainable energy goals, cautioning against overreliance on infrastructure investment and advocating for voluntary cooperation-based strategies that encourage conscientious energy use.**

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

I. Introduction and Research Questions

In a response to climate change, energy systems worldwide are undergoing a fundamental transformation as governments choose to move away from burning fossil fuels and instead aim to generate power from renewable resources, such as wind power, solar power, hydro power and biomass. In Western countries, the installed capacity of wind and solar power is expanding rapidly, meaning more and more decentralized electricity generators are supplying energy into the grid. The EU's most recently published target requires 45% of energy to be generated from renewable sources by 2030 [1]. System operators, which traditionally were occupied with operating and maintaining the grid, are increasingly tasked with coordinating supply and demand and are playing a pivotal role in facilitating the energy transition [2].

The transition is also having a large impact on the the demand side of the grid. The electrification of transport, heating, and cooking are causing rapid growth in electricity demand. The yearly electricity demand from an average household in the Netherlands will double or triple when making the switch to electric heating, cooking, and driving [3]. ENTSO-E projects electricity demand in the EU to grow by around 50% by 2050, despite a fall in overall energy consumption [4]. A significant share of that growth is made up by electric vehicle (EV) charging demand, as the EU aims to ban the sale of all fossil fuel cars by 2035 [5].

EV adoption forms a big threat to the distribution grid, which delivers power to end users. Unlike internal combustion engine (ICE) cars, which need to fuel at dedicated stations, EVs can charge anywhere with a connection to the power grid. When users install a dedicated 3-phase charging station in their house, charging powers often reach 11 kW, over 60% of the average residential connection capacity. For convenience, most users charge the car when coming home after work. If an entire neighborhood does this, while using other appliances as well, the stability of the electricity grid is threatened [6]. Both the charging timing and power are important factors in this issue. They are determined by the charging needs of users, and thus far there is little incentive for users to prevent congestion.

At the same time, falling solar panel prices have spurred the adoption of private solar panel installations for households. In the Netherlands in 2021 about 20% of households had solar panels installed, at an average capacity of 3.7 kWp [7], with more recent installations being larger. Additionally, energy companies and cooperatives are investing in local solar and wind farms. The large level of distributed generation can also lead to congestion, because solar panels within one section of the grid generally receive sunlight at the same time. Existing policy for solar generation also doesn't incentivise users to prevent congestion.

Together, these trends are causing a large change in the usage of local electricity networks. Most of the existing electricity infrastructure has not been designed for these new technologies. Action is needed to prevent issues for grid reliability. Grid operators in the Netherlands already report high congestion in large parts of the country [8], meaning that there are limitations for new connections and expansion of existing connections, as well as limitations for expanding renewable energy sources. Without action, grid limitations will form a significant hurdle for achieving the green energy targets set by the EU.

Many solutions are suggested in academic literature, such as market-based solutions, grid expansion and implementation of local control systems. This thesis approaches the issue from a behavioural perspective, and aims to contribute to the discussion by answering the following question: *what type of behaviour arises from different cooperative rulesets among households sharing limited distribution capacity on a distribution grid, and how can cooperative behaviour be incentivised?*

This thesis combines multiple academic fields of study to answer this question. The concept of strategies emerges from the discipline of game theory, which enables us to mathematically define the households, their behaviour, and their environment. The environment is a simulated residential electricity grid, which connects the players with a common feeder. The behaviour of interacting households in a grid can be simulated and eventually reach an equilibrium using reinforcement learning. The three academic fields and their theoretical overlap are shown in Figure 1.

Over the course of this thesis the following sub-questions will be answered:

- How can we define the rules for a game in which households within an energy community collaborate by load-shifting to reduce peaks on the grid, using tools from game theory?
- How can we describe preferences for each household using a utility function, and what types of (non-)cooperative strategies emerge from these preferences?
- What types of (non-)cooperative behaviour emerge, and what do they imply for the functioning of the grid?
- What do the findings regarding (non-)cooperative behaviour imply for possibly incentive structures to encourage cooperative behaviour?

II. Literature review

A. Grid structure

Since the Single European Act of 1986 the EU has actively passed legislation to stimulate member states to liberalise their energy markets. The aim of this project is vertical

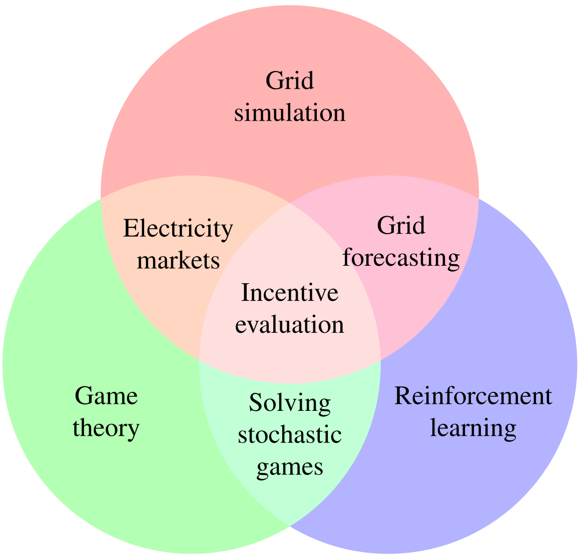


Figure 1: Venn diagram showing the overlapping academic fields addressed in this thesis

and horizontal unbundling of the energy sector and introduce more competition, in markets that were traditionally dominated by (state-owned) monopolies. Ideally, deregulation would drive down prices for end users, boost energy security, and incentivise innovation. Additionally, by aligning market regulation across member states, cross-border trade is enabled. Although according to the Organisation for Economic Co-operation and Development (OECD) most EU member states are highly liberalised [9], Pollitt [10] states that the effect of this liberalisation policy is obfuscated due to the gas and oil price fluctuations and increased taxes and levies on energy as part of pro-environmental policies. One of the downsides of liberalisation was highlighted recently when reduced gas imports from Russia caused energy prices to skyrocket, benefiting large energy corporations at the cost of households [11].

Energy markets in European member states are structured (more or less) in line with the the European harmonised electricity market role model [12] which roughly separates the sector into these elements:

- Regulation
- Generation
- Transmission
- Distribution
- Retail
- Consumption

The European high voltage grid is operated by Transmission System Operators (TSO), which are responsible for the high voltage transmission system within their bidding zone, as well as trading between zones. Most European countries have a single TSO, with Germany being the exception with 4 TSOs.

The transmission system ties together a large number of (medium voltage) distribution grids, which provides grid connection to individual consumers at various voltage levels. The distribution grids are operated by Distribution System Operators (DSOs). The number of DSOs varies from country to country. The Netherlands has 8 DSOs. Although congestion is a significant issue for both TSOs and DSOs, this paper will focus on congestion issues on the distributional level.

According to Hirth and Glismann [13] congestion arises when trade between market parties cannot be transmitted through the grid without violating operating security constraints. So while the responsibility for dealing with congestion lies with the grid operators, it is caused by the trade between generators and consumers, which is facilitated by energy retailers. The markets on which these actors trade operate independently from the grid operators, and the trading platforms do not factor grid constraints into the auction prices. They behave as if there is an infinite grid capacity. At present there is no incentive for a consumer to buy from a generator whose dispatch would produce the least congestion.

B. Congestion on distribution networks

Congestion manifests itself as a threat of current and / or voltage violations. These violations can cause degradation and even damage to infrastructure and appliances, and it is therefore the objective of system operators to prevent violations from occurring.

A fundamental concept at the core of congestion issues is called the **simultaneity factor** (SF), also referred to as the coincidence factor. This was introduced by Rusck [14] who stated that domestic power use (within a given period) is a stochastic independent variable with a normal distribution. This means that, as the number of users increases, the likelihood of them all using a high level of power simultaneously decreases. The simultaneity factor for n users is given by:

$$SF_n = SF_\infty \cdot P_{max,1} \cdot n + (1 - SF_\infty) \cdot P_{max,1} \cdot \sqrt{n}, \quad (1)$$

where $P_{max,1}$ is the peak load of one user, and SF_∞ is the simultaneity factor for infinite users, which is equal to:

$$SF_\infty = \frac{P_{avg,1}}{P_{max,1}}, \quad (2)$$

where $P_{avg,1}$ is the average load of one user. The intuition is that for an infinite number of users, the peaks and valleys in the load would balance out, resulting in a constant load. To get the maximum expected load for the grid, the following equation is used:

$$P_{max,n} = SF_n \cdot P_{max,1} \cdot n. \quad (3)$$

The intuition behind the SF is that users will never all draw (or inject) the maximum amount of current from the grid at the same time, and so DSOs under-size grid infrastructure to save costs. Due to the trends described in Section I, the SF is growing [15], and peak loads are higher and longer than before. A higher SF can lead to power demand that exceeds the design specifications of the infrastructure, which causes a **current violation**. Especially older parts of the distribution network which were designed around smaller SFs are vulnerable to this issue.

Voltage violations, on the other hand, can have multiple causes. Rapid fluctuations in electricity demand or supply that is not immediately matched lead to voltage drops or rises, such as the drop in solar power output caused by a cloud. Secondly, a high level of feed-in of power at the distributional level can cause the local voltage to exceed the established range of operation. Shabbir et al. show that at high levels of PV penetration, voltage violations are more common than current violations due to a lack of reactive power [16].

The maximum penetration level of any technology (such as EV, PV) connected to the grid without causing any violations is also called the hosting capacity (HC) , which can be estimated in a variety of ways, often involving simulation [17]. Yu, Reihs, Wagh, *et al.* [18] showed that suburbs are more prone to congestion caused by EVs, as they generally rely on longer feeders, have larger households, and higher levels of car ownership. A practical experiment in a suburb in the Dutch town of Lochem showed that simultaneous use of EV chargers and high power appliances can melt a fuse in a transformer, leading to a power outage [6].

C. Existing DSO approaches to address congestion

Hirth and Glismann [13] identify two principal options for resolving congestion (at any scale); change the network, or change the demands on the network. Changing the network includes any adjustment, addition, or removal of nodes or edges. Changing the demands of the network can be achieved either competitively, by changing the way markets operate, or cooperatively, by coordinating the demand of users to prevent congestion.

Extensive congestion management is a relatively new task for DSOs, and distribution networks are only weakly monitored [19]. The instruments that DSOs use to address congestion are therefore relatively crude, focusing on adjusting and expanding the network, limiting access to the network, and trading flexibility with large industrial users on dedicated platforms.

DSOs work on optimization and rerouting of network flows in the short term, and expanding and upgrading capacity in the long term. Rerouting of flows often leads to increased energy losses, and precise control requires expensive equipment. Expanding and upgrading is slowed down by financial and human resource limitations faced by TSOs and DSOs. This solution will likely also be less cost-efficient, as the peak-loads which determine the required capacity only occur at low frequencies.

DSOs are legally obliged to provide grid access to anyone who requests it. But in the case of congestion, DSOs can temporarily delay or deny grid access for new users, as well as denying expansion of grid access for existing users, while the physical limitations of the grid are being addressed. This solution provides a partial brake on the growth of distribution demand, although the demand from existing grid users is also growing. The grid is also partially protected against voltage violations from solar power generation through automatic curtailment, which is a feature built into commercial inverters. Similar failsafe mechanisms are being explored for the automatic curtailment of EV charging load [20] to prevent current violations.

The solutions of limiting access and automatic curtailment divide the burden of congestion arbitrarily over (prospective) users. Some users won't notice it at all, but for others it means delaying the electrification of their home, or the cancelling the building of a planned business or public service. User preferences (and political preferences) are not taken into account in this approach, and there is little to no influence on the outcomes.

For areas with large industrial users and/or generators, some DSOs are attempting to find more efficient solutions by coordinating users through a dedicated market platform for flexibility. Bids are requested by the DSO ahead of time, allowing businesses and generators to reschedule operations. Participants are financially compensated for the flexibility provided. Pioneering flexibility markets are discussed in [21], but thus far only few industrial users and generators have participated. DSOs are limited in their ability to cooperate with users by EU law, which states that DSOs must use market instruments for fair procurement of flexibility services [22]. This is to prevent DSOs from using their monopoly position to discriminate users [2].

Thus far, DSOs have managed congestion using crude access denial. Flexibility markets have only shown limited success, and the grid remains vulnerable for current violations. Active research is underway to explore approaches to congestion management that provide more resource efficient solutions.

D. Proposed approaches to address congestion

Bach Andersen, Hu, and Heussen [19] have identified that at different grid scales, different solutions are being proposed in the literature. On the scale of the bidding zone or the TSO, the research focuses on market structures, policies, and regulatory instruments. On the scale of the distribution grid, or even smaller, on the scale of a single feeder, much of the research is instead aimed at finding appropriate control solutions to increase hosting capacity.

This paper focuses on congestion on this distribution scale, close to the end-users, where congestion affects an entire feeder or substation. There are some differences in the nature of congestion compared to the large scale. On a small scale renewable electricity generation is more variable, as clouds and fluctuations in wind speed are localized phenomena [23]. Another difference lies in the roles that users have. On the larger scale, generators and retailers are directly involved in market trading, while on the local scale, users are not involved with the markets at all.

Market based solutions are often seen as an efficient way to to adjust the transmission and distribution demand of market participants, as long as market prices are able to transmit information about the location and timing of congestion. Katz [24] describes how the large time-steps (one hour) and lack of locational information in wholesale market prices provide a barrier to effective market-based congestion management. Aggregated loads (such as EV chargers) could provide significant flexibility, although a minimum portfolio of 500 chargers would be needed to make it economically viable to participate [25]. Verzijlbergh, De Vries, Dijkema, *et al.* [23] suggest that splitting a bidding zone into 'nodes' and reducing the timestep to 5 minutes would allow markets to contribute to alleviating congestion.

Price incentives are a difficult tool to implement for managing congestion. In most proposals the congestion fee is announced per time-step a day ahead, and it does not respond dynamically to changing user behaviour or forecasting errors. Simulations done by Verzijlbergh, De Vries, and Lukszo [26] show that this type of price structure causes users to behave synchronously, shifting and concentrating demand to periods with low prices. In some case this causes peak loads which exceed those of the uncontrolled situation. Different price incentives per user would be required to promote asynchronous behaviour, but this amounts to discrimination. Lastly, the demand for electricity is relatively inelastic because electricity users behave habitually rather than rationally [27], and it is not guaranteed that price incentives can provide a solution.

A different way to implement a price incentive to reduce peak loads is through a capacity subscription tariff for users. At the moment, the yearly network tariff set by DSOs is not progressive, in that users don't necessarily pay more for a larger connection. As described by Hennig, Jonker, Tindemans, *et al.* [28], the capacity subscription tariff, which is already implemented in Belgium, creates a virtual cap on power draw which is lower than the physical limit of the installed fuse. Any peak loads by users that exceed this virtual cap cost extra. Users are therefore incentivised to reduce their peak load to subscribe to the cheapest (and lowest) capacity. Although it is likely that users will reduce their EV charging speed under this tariff structure, even in the rudimentary simulation by Hennig, Jonker, Tindemans, *et al.* [28] congestion starts occurring at an EV penetration rate of 30%. The proposed capacity subscription tariff also offers no solution for synchronous PV generation.

As a different approach for residential areas, many articles suggest a control solution which automatically shifts heavy loads such as EV charging, heating, and appliances, to reduce the SF. Generally, this approach requires extensive communication networks to collect user preferences, calculate a solution that can optimize for multiple objectives, and deliver the schedule back to devices [6]. A randomized scheduling approach is also feasible [29], although it removes any possibility for optimization. Some control approaches also include a battery energy storage system (BESS), which can relieve congestion by absorbing excess supply and feeding in power during peak demand [16]. Although control solutions reduce the need for grid infrastructure upgrades, they require investment into an IT network, which partly undercuts their advantage. Control approaches work well in simulations, but presume the participation of grid users, in voluntarily joining the management program, providing user preferences, and adhering to the proposed schedule. Cooperative user behaviour cannot be guaranteed in a top-down implemented solution.

Existing and proposed solutions for DSOs to manage congestion follow a competitive market mentality. By providing price incentives to users that include both generation and distribution constraints, individual users will sort themselves by willingness to pay for power at any given moment, creating a resource-efficient solution. However, a market solution requires a relatively elastic response from users. Household users are bound by external demands on their schedule, such as work and school times, and social norms, and are not free to rapidly shift their demand based on incentives. Market solutions capitalize on variability, but new habits can only be formed if incentives are expected to remain constant over a longer term. Energy retailers provide a highly predictable price to consumers, while their own business revolves around maximizing profit trading within market constraints. A similar business model would arise if congestion pricing was introduced, which would reduce or remove the strength of the price signal felt by consumers.

Additionally, the proposed solutions increase complexity for grid users, through increased price variability or the need for submitting preferences for energy use. Grid users may not respond well to this change, especially if they are excluded from participating in the formulation of the new rules. In light of the larger energy transition, which has already generated significant resistance in some groups, it is worth exploring alternative approaches to energy management that rely on voluntary participation.

A cooperative solution, rather than a competitive one, could allow users to coordinate their individual demand to achieve a more desirable outcome, and can support the formation of new habits. This paper aims to explore cooperative solutions by using the perspective of a *commons*, which is the term used for resources that are collectively owned and used by a group of people, rather than controlled by the state or a private firm. The literature on commons management is extensive and provides many valuable insights, both for competitive and cooperative solutions. By using the perspective of a commons to break down the behavioural components of congestion, this paper aims to uncover what is needed for a successful congestion management approach.

E. Distribution capacity as a common-pool resource

The distribution capacity provided by a grid can be analyzed as a common-pool resource (CPR). Ostrom [30] provides a foundational theoretical framework for analyzing the management of CPRs, based on extensive research assessing both successful and unsuccessful CPR management case studies. Before Ostrom the prevailing theories in the literature on commons management focused on the need for state intervention or privatization. It was claimed that for any resource system that is accessible by all, over-extraction would eventually lead to the collapse of that resource system. This is also referred to as the *tragedy of the commons*. Hardin [31] famously applies this idea to the issue of overpopulation, claiming that the growing world population would eventually destroy the resource systems that it relies on for survival, and that therefore, intervention was required.

What sets Ostrom's writing apart is her argument that communities can, in some cases, self-organise without the need for centralised control by either a state or a firm, and do so more effectively. She argues that there are many examples of successful self-managed CPRs in the world which cannot be explained by the existing theory of the tragedy of the commons. Although there is not a single solution for all CPRs, certain features of CPRs and the communities that manage them can contribute to or hinder the succes of collective action.

Ostrom also argues that in many cases intervention by the state or privatization has actually caused or exacerbated resource exhaustion and created inefficiencies. Such top-down solutions incentivise competitive behaviour and a focus on short term individual profits rather than long-term sustainability. Private entities benefit from information asymmetry, hiding business practices from each other and regulatory bodies, which in turn leads to speculation and risk-taking. On the other hand, successful management of CPRs benefits from complete information about activities, and resources can be damaged by risk-taking behaviour.

Distribution grids have not yet been covered by literature on CPRs, but the theoretical framework provided by Ostrom provides a broader perspective on the different possible approaches than what exists in the current literature on grid congestion.

Ostrom [30] defines a CPR as a resource system that provides a flow of resource units, which are appropriated by users of the system (appropriators). A CPR is different from a public good, because the resource units are subtractable and limited in supply, such as in the case of limited power distribution capacity. Ostrom identifies two types of issues in every CPR. The issues of *provision*, which relate to the construction, maintenance, and administration of the resource system, and issues of *appropriation*, which relate to the allocation of resources, and the monitoring and enforcement of rules.

The issue of distribution grid congestion is somewhat different from the CPRs that Ostrom focuses on. Before congestion arose as a significant issue, households were seen as passive participants in the grid system, with the issues of provision and appropriation taken care of by the DSO. The DSO can be seen as a state-instituted solution, as it is regulated to address the issues of a CPR. It solves the provision issues by collecting fees from users to take care of the resource system, and it solves appropriation issues by providing limited connection to the grid. Although in today's liberalised electricity sector the DSO is often a private firm, in this framework it is seen as a state solution because the contracts between user and DSO are non-voluntary. Because the DSO holds the monopoly on an essential resource in a given area, it is highly regulated.

Ostrom proposes a division of three levels of nested rulesets for a CPR management institution; *operational rules* are the policies that directly affect the behaviour of appropriators by requiring, permitting, or forbidding certain actions; *collective-choice rules* are the set of rules that is used by appropriators and officials to determine the operational rules; and *constitutional-choice rules* are the set of rules that determines which entities are given authority to make collective-choice rules. To affect any set of rules requires choices to be made in the level above. In the context of a low-voltage electricity grid, the operational rules are generally established by the DSO, while the appropriators can only affect rules through their representatives in government.

Due to the trends described in this paper, appropriation issues have evolved into something that the existing ruleset of the CPR cannot solve. The rising simultaneity of heavy loads indicate that the DSOs have over-allocated connection capacity. Action needs to be taken to change the rules - either by reducing or rotating allocation or by expanding the resource system. This action can be imposed on the appropriators from above, such as by the DSO. However, we have seen the limitations of available approaches in the previous section. It is also possible for the members of the community to organise themselves into an institution to create their own rules. It is not the goal of this thesis to define which of the two is more realistic, but instead to identify possible solutions that have been overlooked in literature. According to Ostrom [32], collective action becomes a real possibility as soon as appropriators themselves are faced with the consequences of system failure, and if they have a joint interest in the long-term stability of the CPR.

A hypothetical group of neighbours sharing a local electricity grid could decide to self-organise to manage congestion issues. In that case, Ostrom [30] suggests 8 design principles to adhere to which she believes lead to long-lasting and effective CPR institutions:

1) Clearly defined boundaries of the CPR, and the possibility to exclude outsiders from joining;
2) Appropriation and provision rules are adapted to local conditions;
3) Those affected by operational rules participate in designing them;
4) Monitoring services are provided by appropriators themselves or by chosen officials accountable to appropriators;
5) Sanctions are gradual and are applied by officials accountable to appropriators;
6) There exist low-cost local arenas for resolving conflicts between stakeholders;

7) Existing governing bodies recognise and facilitate self-organisation;
8) The institution consists of nested enterprises that represent sub-group interests.

Many examples of CPR institutions that adhere to these principles exist. Ostrom highlights examples such as the thousand-year old irrigation rules that are used in the huertas in Catalonia, or rules for managing the commons in rural villages in Japan [30]. In these examples, often a rotating schedule is used for the allocation of resources. Rotation rules are easy to understand and often perceived as fair, because during a random rotation all participants experience both beneficial and less beneficial scheduling. Additionally, a predictable schedule allows users to prepare and adjust their habits to benefit maximally from their allocated appropriation slot.

The above design principles were identified by looking at common patterns in a large number of CPR case studies. The case studies generally described socio-ecological systems (SESs), meaning that the CPRs studied are natural resources. The electricity grid is not a natural resource, but a technical one, meaning the system being analysed is a socio-technical system (STS). Acosta, Ortega, Bunsen, *et al.* [33] provide a precedent for the application of applying SES theory to STSs. Their analysis suggest that the above design principles are appropriate for the design of an integrated community energy system, in which a community self-organises to provide local energy needs. More recently Ostrom's design principles were applied by Eslamizadeh, Ghorbani, and Weijnen [34] to an agent-based simulation which was used to test the institutional designs for an industrial community energy system in Iran.

However, both studies fail to address a critical difference between SESs and STSs: the natural resource in an SES is (generally) **buffered**, meaning that the availability of the resource units varies over time, and is affected by in- and out-flows. A technical resource in an STS such as distribution capacity generates resource units with no lifespan - these units must be appropriated immediately and cannot be stored. The appropriation limit of the technical resource is capped at a constant value. Stability of an SES is threatened by over-exploitation over an extended period of time, while stability of an STS is threatened by over-exploitation in one instant of time, caused by the misaligned behaviour of appropriators rather than excess total demand.

A key contribution of this thesis will be the application of Ostrom's design principles to an STS in which the generation and appropriation of resources is time-constrained. More characteristics of such systems will be further elaborated in the next section.

F. Managing commons: learning from traffic congestion

In subsection II-B it was shown that the fundamental cause of distribution grid congestion is the rising simultaneity factor. Especially suburbs are prone to congestion, as the daily schedules of suburban households are very similar, leading to synchronized power profiles where users heat, cook, and charge at the same time. Dense urban neighborhoods are less prone to congestion, because they benefit from a diversity of users in various ways: diversity of mobility means users rely less on private vehicles [18]; diversity of lifestyle means residential users are less likely to use power at the same time [35]; and diversity of zoning means that commercial (and sometimes industrial) functions are mixed into the distribution grid, creating a more even load throughout the day and night.

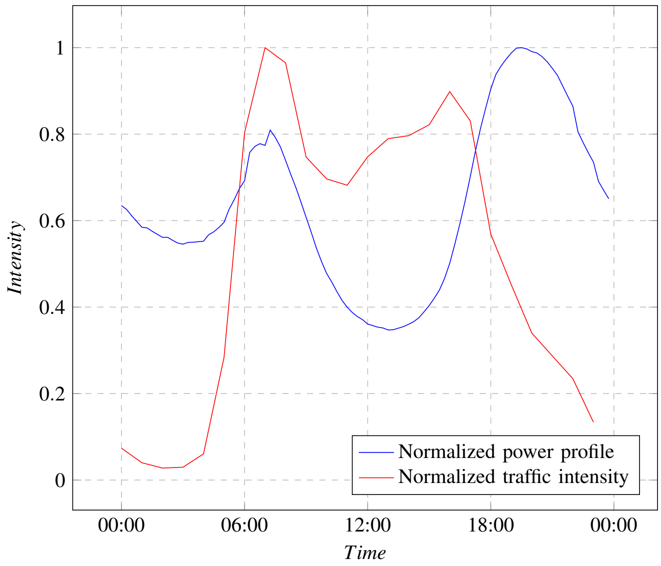


Figure 2: Graph showing the normalized average daily demand of a typical Dutch household electricity connection in 2023 ([36]) and the average traffic intensity northbound on the A29 highway on weekdays [37]

For practical insights and cautionary tales, a comparison can be drawn between congestion on the electricity net and congestion on the road. Both distribution and transport capacity are a flow of resource units that is generated by an STS consisting of a network of infrastructure elements. These resource units are time-constrained - they have to be used as they are generated. In both fields congestion of the network has a non-linear relationship with demand - only a peak in demand (rush hour) that crosses the available capacity of the network leads to issues, but the issues are significant and felt by all users on the network. Congestion even follows a similar profile; Figure 2 shows the average usage intensity of both a typical highway and a typical residential power connection over the course of a day (averaged out through the year). There are clearly two peaks in demand, one in the morning, as people prepare and travel to work, and one in the evening, as people travel home and spend their evenings there.

The existing literature on traffic congestion is mature, extensive, and provides many insights that can be applied in the field of grid congestion. The largest insight is that, despite years of research and experimentation, traffic congestion has not yet been solved.

Infrastructure expansion seems like an obvious solution for traffic congestion. However, in practice the issue is subject to a couple of counter-intuitive phenomena. Braess' paradox [38], published in 1968, states that infrastructure expansion can worsen congestion, as the expanded road capacity induces more demand and / or shifts congestion to a point that is less able to handle congestion. The phenomenon occurs in power networks as well. Witthaut and Timme [39] show that expansion of the electricity grid can actually worsen congestion in some cases. For example; only about half of charging sessions in the Netherlands take place at home [40]. The expansion of the distribution electricity grid could allow users to shift their charging demand from public / office chargers to home chargers, inducing a high synchronous load which could cause congestion that is worse than the congestion pre-expansion.

A related concept from this field, the Downs-Thompson paradox, states that the travel time by car is equal to the travel time of an equivalent journey by public transport (subject to available capacity) [41]. That is to say, users will avoid congestion if they can save time by taking the train. According to this paradox, providing a viable, high capacity alternative is therefore seen as the only solution to traffic congestion in a situation where the schedules of users aren't up for debate. A literature review by Kuss and Nicholas [42] shows that control solutions such as congestion charging can be effective, but the case studies only consider cities where public transport is a viable alternative.

The Downs-Thompson paradox describes the interplay between different *resource vectors* that provide similar services. In transport, congestion can be alleviated by using public transport that has higher capacity but, to achieve this capacity, has a lower flexibility in terms of destinations. In energy systems, alternative vectors to electricity exist in the forms of liquid fuel (for transport) and natural gas (for heating and cooking), which provide very high capacity but are only suitable for a single purpose (see also Table I). In the past, a user that, due to congestion, could not obtain an adequate grid connection for their electricity demand, had the option of switching part of their demand to a different vector, such as heating through natural gas. As a result of the energy transition, the access to energy vectors other than electricity is reduced, causing grid congestion issues to skyrocket. It is analogous to the traffic congestion that would be caused by removing all public transport from a city.

Without viable alternatives to electricity, neither infrastructure expansion nor control solutions can fully alleviate congestion, and attempting to do so would be a highly ineffective use of resources. Transmission and distribution capacity are and will remain a scarce resource, and due to induced demand, peak loads will always grow to meet the limits of the infrastructure.

This thesis suggests that both traffic and grid congestion are part of a class of problems which Hardin [31] describes as "no technical solution problems", meaning that it cannot be solved through natural sciences alone, but that it demands change in social values and behaviour. If the world is to fully embrace electricity as its only energy vector, individuals will need to charge, cook, and heat at different times from each other, which is counter to existing habits. Another approach to the energy transition would be to make high power vectors available when needed, such as supplementing green hydrogen gas or biomass to support heat pumps in winter.

One approach to scrambling behaviour could be to coordinate days in which users work from home. Working from home reduces peak load for traffic, as well as for the grid, as people working from home use appliances at different times than people commuting. In an optimal solution, work-from-home days are evenly spread throughout the week, uniformly reducing demand. Unfortunately, work-from-home days are picked without coordination and individual preferences have led to sub-optimal division: in the Netherlands Mondays, Wednesdays, and Fridays are much more popular work-from-home days. High travel demand on Tuesdays and Thursdays is therefore still causing major congestion [45]. When each user individually chooses their work-from-home days the outcome is highly sub-optimal, suggesting that coordinated action is the only possible approach.

The following conclusions are drawn from the comparison between grid and traffic congestion:

- Newly created space on a resource-providing network induces additional demand;
- The availability of alternative resource vectors can effectively reduce peak demand;
- Incentives are subject to interaction with alternative resource vectors and social norms.

Researchers should take care to recommend infrastructure or control solutions in which the demand is assumed to be fixed. The demand of households, when restricted, will likely behave in complex ways, and cannot be expected to remain fixed when a control solution is implemented. This thesis aims to contribute to the pool of solutions by exploring how effective a coordinated behaviour change of users could be.

G. Game theory

Game theory is a field of study that looks at the interaction between agents in situations where their joint actions determine the overall outcome. Game theory thereby attempts to model a framework that is used for decision-making. It is commonly applied in the study of commons-management, where it is helpful in identifying free-rider behaviour. Game theory is also widely used to study energy systems, both for deepening our understanding of its functioning and for the development of control approaches. This thesis uses the book by Fudenberg and Tirole [46] as a reference for game-theoretic principles.

The liberalisation of the energy markets discussed in subsection II-A has led to interest in game theory in the literature. The liberalisation meant unbundling the utility companies into

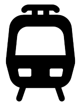
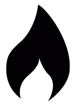
Maximum throughput [43]	**Transport vectors**	**Common characteristics**	**Energy vectors**	**Average connected capacity** [44]
2,000 pass./hour/lane	 Private car	• Prone to congestion • Highly flexible	 Electricity	17 kW
80,000 pass./hour/lane	 Rail	• High capacity • Single task	 Gas	60 kW

Table I: Highlighting similarities in vectors for mobility and energy.

the individual elements described in subsection II-A. Pre-liberalisation, state-run utility companies had control over most or all of the energy generation, transmission, and distribution resources. Now, generators and suppliers have to compete for profit. TSOs and DSOs hold regional monopolies, but are able to extract profit by lowering their cost of operation. The market is split up into a large number of agents, each motivated by their own profit.

It was thought that introducing competition in the energy market would improve efficiency and reduce costs for end users. Regulatory instruments are needed to ensure that the power grid is stable, and to ensure fair competition. Since there are many agents in this market whose behaviour cannot be planned, the government needs tools to analyse their interaction and predict their response to regulation. Game theory provides powerful tools for this type of analysis.

The utility theorem of Von Neumann and Morgenstern [47], published in 1944, forms the basis of modern game theory. It theorizes that rational actors will make decisions by maximizing the expected value of a reward function R. The existence of such a function implies that game-like scenarios, which occur in many fields including economics and biology, can be mathematically solved to find the optimal strategy for any player. A player uses a 'mixed' strategy σ_i if they choose an action from an action set according to a probability distribution. If only one action is feasible (and therefore has a probability of 1), then the strategy is said to be 'pure' (s_i).

Nash [48] elaborated on this by introducing the concept of the Nash-equilibrium. This is a type of equilibrium in which no player has an incentive to (unilaterally) deviate from their current strategy, given that the strategy of the other player is known and assumed to be fixed. An example of such an equilibrium will follow. It has been proven that a mixed strategy Nash equilibrium exists for any finite game (finite implying a finite number of rounds, players, and actions). The computation of the Nash equilibrium is NP-complete, meaning its computation is non-trivial.

To illustrate the utility of game theory for analyzing the appropriation of common distribution capacity, let us consider a game described by the tuple (N, A, R), where:

- N is the number of players;
- A_i represents the action set for player i;
- r_i represents the reward function for player i, which maps a reward to each available action.

2 players share a power grid of which we will set the total distribution capacity equal to 1. In this simplified scenario, each players action set consists of 2 options, either using a low amount of power (L), or a high amount of power (H). The game is static, meaning the players must choose simultaneously. The payoffs are related to the amount of distribution capacity consumed. The game is illustrated in normal form in Table II. The payoffs are arbitrarily determined in a way that allows for one, but not both players to play H, and represent available distribution capacity.

If both players choose L, for example by actively spreading out their demand throughout the day, the payoff for each player is 0.3. In this case 60% of the available distribution capacity is used. If one of the players chooses to increase their demand by choosing H, the use of distribution capacity rises to 90%. However, if both players choose H, the demand exceeds the available capacity, the grid fails, and players receive payoff X. This type of game is a variant of the *coordination game*, also called the *game of chicken* [46]. This game is a *general-sum* game, meaning that one player's gain does not not necessarily mean another player's loss, as it does in a *zero-sum* game.

As described above, the Nash equilibrium describes the strategies for which neither player can obtain a higher payoff by changing their strategy. In a single-stage game (which is played only once), only pure strategies can be played. Let's consider the strategy profile (L, L), meaning that both players play L. Each player can increase their payoff from 0.3 to 0.6 by changing their strategy to H. However, if the strategy profile is either (H, L), or (L, H), then neither player can increase their payoff. Switching from H to L reduces the payoff from 0.6 to 0.3, while switching from L to H will lead to a load violation, causing grid failure, and giving payoff X (which is assumed to be 0 or lower). Therefore (H, L) and (L, H) are Nash equilibria.

If we consider that this game is played many times, then a mixed strategy could increase the payoff for each player. One way to solve for the Nash equilibrium is by making assumptions about the actions that are given a positive probability in each players strategy (these actions form the *support*). If we assume that all actions are part of the strategy, then each player must set their strategy in such a way that the other player is indifferent about their choice, since all actions have the same expected return. If this wasn't the case, then they playing a pure strategy would give a better result.

p and q respectively indicate the probability of playing L for player 1 and 2. p must be set in such a way that the expected payoffs for player 2 are equal to each other:

$$E_2(L) = E_2(H) \tag{4}$$
$$0.3p + 0.3(1-p) = 0.6p + X(1-p) \tag{5}$$
$$0.3 = (0.6-X)p + X \tag{6}$$
$$p = \frac{0.3-X}{0.6-X}. \tag{7}$$

$p = q$, since the game is symmetric. The mixed-strategy equilibrium strategy depends on the penalty X. If X is 0, the Nash-equilibrium strategy is to play L and H each with a probability of 0.5. That way, each player's opponent is indifferent over both choices, as the expected reward (the probability multiplied by the payoff) of each choice is equal to 0.3. In this case, the chance of the game resulting in grid failure is 0.25.

Assuming that grid failure has large negative consequences for each player, due to cost and time for restoring the power, the value of X can be lowered to make this outcome less desirable. As X decreases, the players are less and less likely to choose H, as the risk of grid failure is too daunting. Figure 3 shows the relationship between the penalty and the chance of not cooperating in the mixed strategy Nash equilibrium.

As long as the game is played competitively, the total payoff is limited to 0.3, and the risk of grid failure continues to exist. However, players could increase their reward by coordinating. If the player agree on a rotation schedule which defines when each player is allowed to draw more power, like a traffic light at an intersection, their average expected reward over time is equal to 0.45. This is also known as a correlated equilibrium, since the actions of the players are now correlated to each other.

The above game is a static game. In a real power grid, users rarely make decisions at the exact same time. In this case, it makes more sense to consider a dynamic game, where choices are sequential. This game can be formally described using the tuple:

- N is the number of players;
- H is the set of choice nodes;
- Z is the set of terminal nodes;
- χ is the action function which maps actions to choice nodes;
- ρ is the player function which maps the active player to a node;
- σ is the successor function which maps a choice node and action to a new node, either choice or terminal;
- r_i represents the reward function which maps a reward for player i to a terminal node.

An example of a dynamic game in extensive form is shown in Figure 4. Player 1 makes a decision first, and is free to choose L or H. Then, player 2 can choose between L or H. Each choice node is referred to as a *subgame*. In a game of *perfect information*, player 2 is aware of the first players choice. If player 2 tries to maximize their payoff then they will always pick the choice with the higher return - mixed strategies are not helpful in this case. A strategy that includes a choice in a subgame for a lower than maximum payoff, is said to be not subgame perfect, as the strategy contains a non-credible threat. The equilibrium for this game is $(0.6, 0.3)$ - because player 1 chooses first, they are always able to get the higher payoff.

The game described is a dynamic game of perfect information. In a game of imperfect information, players have limited knowledge about the choices of other players. Choice nodes that are indistinguishable from each other are said to be in an equivalence class. If no information is available about other players' choices, then the resulting game is identical to a static game.

Next to dynamic games where players play sequentially, one can also consider an iterated (repeated) game where players play the same game a number of times. In that case, players attempt to maximize their total reward over all games rather than in one game. Since a player's action may influence the other player's future actions, the employed strategies become much more complex. Axelrod organised multiple tournaments to research possible strategies for playing an iterated prisoners' dilemma, in which submitted strategies were pitched against each other in successive rounds to determine the strategy that scores highest overall [49] [50]. Axelrod identified the following qualities in successful strategies:

- Niceness: not being the first to defect;
- Forgiveness: ability to reestablish cooperation after the opponent has defected once;
- Provocability: immediate retaliation if the opponent defects.

The strategy that won the first two tournaments was called tit-for-tat, in which the player simply copied the previous move of the other player, and it possesses all of the above qualities. Note that provocability and forgiveness correspond to Ostrom's design principles 5 and 6, which call for sanctioning any infraction, but only lightly, so that conflict between players can be resolved quickly and collaboration reestablished.

		Player 2	
		L	H
Player 1	L	0.3, 0.3	0.3, 0.6
	H	0.6, 0.3	X, X

Table II: Normal form representation of the distribution capacity game

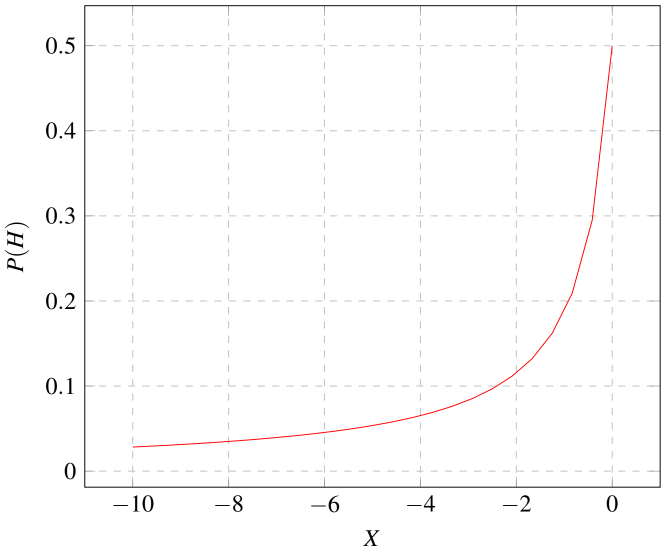


Figure 3: Graph showing risk-avoidant behaviour as X decreases

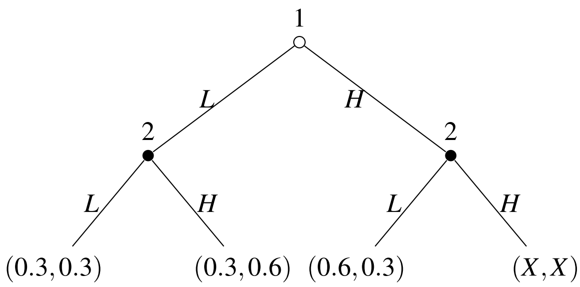


Figure 4: Extensive form representation of the distribution capacity game

Axelrod's approach to discovering optimal strategies was experimental rather than mathematical. The submissions were often based on heuristics, sometimes attempting to predict the behaviour of the other player, and using generalized knowledge about the game to adjust their actions. These heuristics-based approaches led to widely varying outcomes when pitched against each other. The trade-offs between mathematical and experimental evaluation of strategies will be further discussed in subsection II-J.

In later works, while analzying the fundamental drivers of international cooperation, Axelrod comes to similar conclusions as Ostrom about the need for long-term commitment of players to enable collaboration, the need for transparency about the actions of others, and the importance of institutional context [51]. This is particularly notable because the conclusions from Axelrod were drawn from theoretical game theory, while Ostrom distills her conclusions from case studies.

Regarding the application of game theory to the field of energy systems, Abapour, Nazari-Heris, Mohammadi-Ivatloo, *et al.* [52] have done a survey of papers that analyse energy systems using game theory between 2015 and 2017. They classified papers according to the type of game setup. The overview shows that many different classes of games find extensive use in literature related to the energy grid, but that its core applications are economic analyses and optimization problems. Game theory is rarely used to analyze incentives to voluntarily participate in energy management approaches.

H. Stochastic games

The games described above are single-stage games, meaning that players aim to maximize their payoff over a single period. For congestion management, it makes more sense to consider a repeated game, in which players aim to maximize their payoff over the course of a number of stages. This more closely represents the use of an electricity grid, as it allows for strategies in which players shift their loads to intervals when other players are not using much power. Specifically, the model of a stochastic game may be appropriate, which was first introduced by Shapley [53], and which allows for the representation of the environment as a state.

Stochastic games are a multiplayer extension of the concept of a Markov Decision Process (MDP). A Markov Decision Process is described by the tuple:

- $S_1, S_2, ...S_n$ is a set of states;
- A represents the action set;
- T is a transition function which determines the probability of moving from one state to the next for a given action;
- R is the reward function which determines the reward for a given action in a certain state.

A visual representation of a simple MDP is displayed in Figure 5. This process mimics the games previously described, but only has one player and introduces a stochastic element. The player starts at state S_1, representing normal grid function

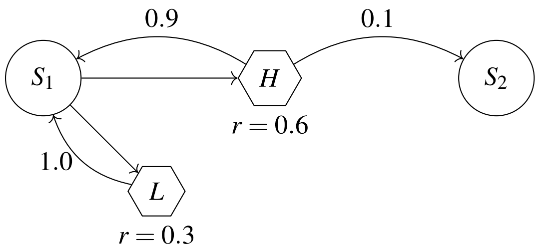


Figure 5: Markov Decision Process representing a single-player grid congestion game

The player can choose between actions H and L. Each action has an immediate reward r associated with it, as well as a probability distribution determining the next state. If the player picks action H, the chance of maintaining a stable grid (S_1) is 0.9, while the chance of grid failure (represented by S_2) is 0.1. The player can also pick action L, which has a lower reward but no chance of grid failure.

A policy π can be used to prescribe a given action for each state. The optimal policy is one that maximizes the agent's total reward over time. Generally, future rewards are discounted with a discount factor γ, to represent that immediate return is valued over a future return.

A stochastic game extends the concept of an MDP by introducing multiple players. The transition function and reward functions are now determined by the joint action space $A = A_1 \times \cdots \times A_n$. Depending on the number of players, the influence of a single player on their own reward and on determining the next state is reduced. Stochastic games are similar to repeated games, except that instead of playing the same game each turn, the stage game is determined by the state, which is in turn determined by the results of the previous game.

It has been proven that Nash equilibria solutions exist for stochastic games [54], in which each player has determined a policy π from which deviating would cause a reduction in expected reward. However, determining this equilibrium is highly complex due to the interaction of players' behaviour. This will be further discussed in subsection II-J.

In this section, different classes of games have been considered to formally define the type of game that users of the electricity grid are playing to manage congestion. Before discussing potential ways of determining an equilibrium, first potential ways of defining the reward function R must be discussed.

I. Reward generated by congestion management

It is required to assume that players are rational to be able to identify equilibria in any game. Rationality implies that players will always choose an action, strategy, or policy that leads to the highest individual reward. This assumption does not give a clear answer on how reward should be modeled; in the case of a low-voltage electricity grid, the most logical reward is the utility each player obtains from using the electricity grid. However, the electricity grid is part of a more complex socio-technical system, in which individual habits and social practices have a large influence on determining behaviour, as exemplified by the road users described in subsection II-F. Ostrom [30] claims that in stable CPRs, sanctions are generally small, because the social cost of infractions is considered high, assuming that the CPR is managed by a tight-knit community. Loss of reputation may reduce a player's influence determining the rules of the CPR at a collective or constitutional level. It is difficult to quantify the social cost in a way that can be compared to the measurable reward from grid use.

Ostrom [30] also mentions that appropriators are generally driven to compliance because they believe that the management institution is successfully preserving the CPR, while simultaneously allowing individuals to represent their own interest. The continued existence of the institution is a reward in itself, which, just like social cost, cannot easily be quantified. In most repeated games, future rewards are included at a discounted rate, which can incentive actions that keep the grid stable over actions that lead to grid failure. But for the example of the grid, grid failure should be distinguished from institutional failure. Exhaustion of the CPR (the local grid) causes a grid failure, meaning that actions which are more likely to lead to exhaustion are punished. But the fact that repeated grid failures may lead to the failure of the institution, and consequently, will likely lead to the implementation of a new management system of which the rewards may be lower, such as the expansion of infrastructure by the DSO, cannot be quantified in terms of reward for individual actions.

Therefore, it is suggested by Ostrom [30] that to be able to do any meaningful analysis, an analyst should focus on one level of choice while keeping the corresponding ruleset fixed. This thesis aims to evaluate different operational rules that could be adopted by either the DSO or a self-governing group of grid users. Therefore, the operational rules should remain fixed, giving players free choice of actions and strategies for appropriation within the given ruleset. Rewards and punishments related to the constitutional or collective-choice level, such as institutional failure or loss of influence in setting policy must not be included. The possible rewards of congestion management will be evaluated through this perspective.

The most direct reward for grid users is the ability to draw power from (or inject power into) the grid. The amount of power also matters, as different devices can be switched on or off depending on the available capacity.

An important choice to consider is whether to include the financial benefits of any power usage strategy. There are some financial benefits for individuals that participate in load shifting or congestion relief, but these are likely to be low and subject to risk. In 2022 energy products made up on average 10% of household spending in the EU [11], and electricity use accounts for about a third of that. Using a contract with hourly rates, households can save a fraction on energy cost.

Gains are higher for households with high electricity use, such as households with EVs or heat pumps. Since congestion is not (and probably will not) be accurately factored into market prices, a cost-minimizing approach is not guaranteed to benefit congestion [26]. Individuals can participate in aggregated EV charging schemes, where portfolio managers trade flexibility on energy markets, and receive some compensation.

Generally, users have shown to be risk averse with a preference for fixed prices over dynamic prices [24], although this is subject to price stability and public perception [23].

Another source of financial benefits could be network tariffs. These prices are generally fixed per year by DSOs. Users could save on network costs by downgrading their network connection. However, network tariffs are not generally progressive, and there are not many options for users. Even in the case of a capacity subscription tariff, as discussed in subsection II-D, downgrading would cause a permanently reduction in utility, which is a blunt solution for addressing congestion which occurs infrequently.

There are other financial benefits to participating, but these are not distributed proportional to the individual input. Good congestion management would yield:

- Reduction in curtailed solar power (benefiting the owners of solar panels);
- Space on the grid for expanding or adding connections (benefiting the users of those connections, as well as the DSO through additional revenue from network tariffs);
- Reduced need for investment into the grid (benefiting the DSO and institutions that fund these investments);
- Other local grids with worse congestion may be prioritized for investment (benefiting the users of that local grid);
- Reduced impact of infrastructure upgrades such as cable replacement and installation of transformer cabins (benefiting the neighborhood).

This raises the issue of fairness. The active participation of one household generates benefits for many. However, many of these issues don't occur at the operational level, and will be excluded from the current analysis.

Although we can state that the behaviour of grid users is rational, that does not mean that it is exclusively driven by financial interests. In studies tracking the promotion of sustainable behaviour, participants prioritize environmental reasons over financial reasons, especially when financial incentives are small [55] [56]. Behaviour can also be influenced by gamification [42], such as the ranking of users, which creates some additional social pressure to participate.

Mathematically modelling the reward function is a highly complex task. By making assumptions and fixing the ruleset, it is possible to evaluate outcomes for a certain congestion management policy. Research suggests that financial benefits do not adequately cover the reward. For a congestion management approach in which a limited amount of distribution capacity is available, an appropriate way to measure reward may be the obtained power from the grid. The possibility of using financial cost for the reward function will be revisited in the discussion in section V.

J. Finding equilibrium beliefs

The traditional approach to 'solving' scenarios in game theory is to use mathematical expressions which have been proven to accurately determine or converge to one or more (Nash) equilibrium points. Bowling and Veloso [54] highlighted a number of algorithms that can be used for solving multi-player stochastic games. Many of these approaches have limitations, such as requiring certain game rules, limiting the behaviour of opponents, or requiring the existence of one unique equilibrium.

In a complex stochastic game with many heterogeneous players and high dimensionality, such as the presently considered grid congestion game, theoretical foundation for determining a Nash equilibrium is lacking. Determining an equilibrium is difficult because of a number of reasons. Firstly, the game is highly dynamic. Player X is attempting to maximize their own reward based given what it believes about the actions of others. But the actions of others are in turn partly contingent on how player X behaves. This creates feedback loops of changing behaviour that are difficult to predict. Secondly, the rewards in the game are non-continuous, fully rewarding players for their actions until the load limit is reached, after which rewards are 0. Lastly, stochastic variables introduce noise into the behaviour of players and the rewards.

A mathematically sound equilibrium is difficult to achieve for this type of game. However, from a behavioural analysis perspective, the loops of adjusting behaviour are more valuable in our understanding of real-world dynamics than are calculated outcomes. In practice, players will act according to a set of subjective beliefs rather than through a calculation of expected reward, or as Axelrod and Keohane [51] described it: *"Beliefs, not realities, governed conduct."*

The set of beliefs about the system itself and about the actions of others, inform the player of an optimal policy. Beliefs are updated over time in response to experiences that deviate from expected outcomes.

Ostrom claims that two core beliefs are essential for maintaining quasi-voluntary compliance from all players; that (1) other players are (mostly) compliant, and (2) that in the long term the expected net benefits of collective compliance exceed the expected benefits of any individual strategies [30]. These beliefs need to be regularly reinforced. Player infractions form an important way of updating these beliefs. For a given appropriation ruleset, and a given set of beliefs, a player's optimal policy can include compliance as well as infractions. Players can be expected to break the rules for one of three reasons; (1) because the circumstances demand it, such as when a household charges their car when they are not allowed to because they have to make an emergency trip; (2) because players are testing whether or not they can increase their reward through infractions; or (3) because of player error.

The offending player learns about the available reward for breaking the rules and the chance of being detected. The other players learn, by being notified of an infraction, about the functioning of the monitoring and enforcement system. If no infractions were reported, then players would have no evidence to base their beliefs on. Because player infractions are vital in maintaining compliance, Ostrom suggests that sanctions should be graduated, starting small, and rely more heavily on the social cost of a detected infraction [30].

In the congestion management environment, different operational rulesets could exist to prevent violations. The sanction for rule-breaking can manifest itself as grid failure, and in the case of active monitoring, a fine. In case the local grid is managed by the community, the social cost of an infraction could be relatively high.

One method for simulating the evolution of the beliefs about the expected rewards of either charging or waiting is called **reinforcement learning**. This is a type of machine learning where simulated entities play the same game many different times, adjusting their behaviour towards taking actions with the highest possible reward. Bowling and Veloso [54] describes how reinforcement learning can be used in game theory to find a (local) equilibrium. The equilibrium represents the point where players beliefs, and therefore their behaviours, are stable.

The reinforcement learning approach for determining the optimal policy shows similarities with Axelrod's tournaments. The game is iterated to determine what behaviour leads to the highest reward in the long term. But while the strategies submitted to Axelrod's tournaments were static, the reinforcement learning approach allows player to dynamically update their beliefs and behaviour based on experience.

As mentioned in subsection II-G, simulations are often used to predict grid performance outcomes. This thesis aims to combine grid simulation, game theory, and reinforcement learning to evaluate the resulting beliefs for a number of operational rulesets which arise from congestion management approaches discussed in subsection II-D. In doing so, dominant policies within the ruleset can be identified, and the rules can be adjusted for more desirable outcomes.

K. Synthesis

Examination of the literature across diverse academic fields has provided valuable insights into the fundamental causes of and possible solutions for congestion within the low-voltage electricity grid. A brief policy analysis has shown how liberalisation of energy markets in the EU has led to an electricity market structure in which congestion is rapidly increasing, but in which end-users aren't incentivised to limit congestion. Existing and proposed solutions in literature rely heavily on market based approaches, but are limited by the low spatial resolution and temporal sluggishness of existing market structures. For local distribution grids, control solutions are suggested to prevent congestion, but no solution is provided for the resource constraints and legal limitations faced by such solutions. A parallel is drawn with traffic congestion, which occurs in a similar type of network-based STS. Despite decades of research, traffic congestion has not been solved, and the only antidote that has been proven effective is to take people off of the road altogether.

By looking at the local distribution grid through the perspective of commons management, another approach to addressing congestion is brought into view. Cooperative rather than competitive behaviour, is shown to be effective for communities sustainably managing an *ecological* resource under certain conditions. Commons theory provides valuable insight for successful management of resources by a community. Although previous applications of commons theory to the management of a common energy system exist, this thesis proposes that it can also be applied specifically to the management of the distribution grid. Due to the unbuffered nature of the resource, the incentives that are present in the system need to be carefully examined.

Game theory is commonly applied to analyse whether players sharing a common resource are incentivised to either compete or cooperate. Game theory can also be used to generate suggestions for altering the ruleset to improve outcomes for all players. The complex nature of a households sharing a distribution grid can be modeled using a stochastic game. An iterated simulation of this game allows for the evaluation of incentives that are visible to each player.

The following chapter proposes a methodology which applies game theory to a low-voltage grid, with the aim of examining incentives to manage congestion in various (non)cooperative scenarios.

III. Methods

The goal of this research is to evaluate incentives for congestion management in a residential electricity grid. To do so, a methodology is proposed that simulates a grid where players aim to maximize their reward following a set of rules derived from game theory. Reinforcement learning is used to allow players to adjust their strategy in an attempt to maximize their reward. The following subsections describe the methodology in detail.

In the following game theory subsection, the rules of the grid congestion game are described, as well as the possible levels of player information, which are represented as states. In the grid simulation subsection, the design of the residential grid and the methodology for evaluating the grid state is described. Next, the algorithm for reinforcement learning is introduced. Finally, a set of hypotheses is given which can tested using the proposed methodology.

A. Game theory

Using game theory, the situation in which players share a local distribution grid can be described using the following statements:

- A set of N players play a stochastic game which lasts for T stages;
- Each stage game $t \in T$ is associated with a state $s_t \in S$;
- The action set of each player is $A_{n \in N} = (wait, charge)$, and A is the joint action space $A_1 \times \cdots \times A_n$;

- After each action, each player is rewarded with a utility $r_{t,n}$ which describes the extent to which they are able to fulfill their energy demand;
- The joint action space A_t for s_t partly determines the state of the next stage s_{t+1}, while another part is determined by independent variables;
- The maximum available utility is limited by the distribution capacity of the grid $\sum_N r_{t,n} \leq R_{max}$;
- If the maximum available utility is exceeded, then the utility for all players is 0 for this stage game;
- Each players n must devise a policy $\pi_n : S \rightarrow P(A_n)$, mapping each state to probability distribution over the action set, which maximizes their individual utility.

The length of a stage game, which can be considered one round, is arbitrary, but has an impact on simulation complexity. In this thesis one stage game describes the decisions of all players for 1 minute.

The above statements do not yet fully specify the game that is being played; the set of possible states S still needs to be specified. The state describes the information available to the players; for example, a state could include information about local voltage, grid load, time of day, and any additional constraints that have been created for grid congestion management. Players aim to select the action with the highest expected reward for any given state. We have seen that many different congestion management setups are discussed in literature, which correspond to a variety of game setups:

1) Uncontrolled,
2) Voltage informed,
3) Load informed,
4) Rotation,
5) GridShield.

In an **(1) uncontrolled** scenario, players have no information about the actions taken by others or about the state of the grid. The decision of using any device within a round can be seen as an action, but practically, the devices whose use matters are high-load devices such as EV chargers. The results from Nijenhuis [57] is used as reference for the uncontrolled scenario.

There are many conceivable scenarios in which some form of congestion management is applied. In a player informed scenario, the players receive information about the state of the grid. This information is highly valuable to players, as they are able to tailor their actions to the available grid capacity, thereby preventing power outages. The most rudimentary grid information that can be obtained by players is the **(2) grid voltage**. Players could measure this themselves without input needed from the DSO. A high load causes the grid voltage to sag. However, this sag is more pronounced towards the end of the grid (as illustrated in Figure 6), meaning that players only get information about the power use of players between themselves and the transformer. This creates asymmetric information for the players. In reality, voltage can also vary because of events happening in the larger medium-voltage grid that the residential grid connects to. This is not simulated in this thesis, but overall it can be said that local voltage is a poor proxy for grid load.

To obtain accurate information about the **(3) grid load**, the DSO could monitor the loading of the cable at the transformer and relay this information to players. Assuming that an EV charger draws 11 kW when switched on, players know that they cannot charge if the remaining available grid capacity is less then 11 kW. Charging would cause grid failure, which reduces a players reward more than waiting would, as they can no longer draw the (much smaller) base load for their home. However, even when informed of the grid load, grid failure remains a possibility for 2 reasons; firstly, although the base load is assumed to be fixed and not included in the action set, electric cooking could cause a load increase of a couple of kilowatts which may be enough to trigger a grid failure. Secondly, it is possible that in rare cases players make decisions simultaneously. If it is assumed that there will be some delay between the checking of the grid data and switching on a charger, then during this delay, the grid load could have increased due to another player starting to charge. An additional charger can then lead to grid failure.

The above scenarios describe a *competitive* approach, where players individually try to maximize their utility. However, individual decision making, albeit well informed, does not rule out the chance of grid failure, and can lead to unfair and/or undesirable outcomes. A competitive approach favors players that start charging earlier, which could incentivise players to try to get home earlier. To avoid these types of side-effects, a number of coordinated scenarios is considered in which some form of organisation (either internal or external to the community), sets appropriation rules for grid capacity, prescribing allocated amounts or time-slots. Additionally, the organisation can provide rules for monitoring and sanctioning of players. Although the ruleset would be designed to maximize the common good, individual players may still be incentivised to break these rules. Whether or not a player does this depends on whether they have grid information available, what the likelihood is of being sanctioned, and how heavy the sanction is for deviating.

One of the simplest rulesets for sharing grid capacity would be to create a **(4) rotation** system for allocating distribution capacity to users. Players are assigned a certain interval in time in which they are allowed to charge. Rotation systems are simple to understand and have been proven highly effective in the management of CPRs.

Another type of coordinated approach is the use of a control system for the management of loads, such as discussed in subsection II-D. One mentioned control system is **(5) GridShield** [20], where a network of EV chargers is programmed to automatically and uniformly curtail the load if the grid is nearing capacity. The key difference with a rotation system is that charging remains possible at all times, albeit at reduced power. This promising technology faces some hurdles in its implementation. Grid codes for load curtailment, such as exist for solar power, do not yet

Variable	Value	Unit
R	0.642	Ω/km
X	0.083	Ω/km
C	210	nF
I_{max}	142	A
q	50	mm^2

Table III: Characteristics of the NAYY 4x50 SE cable type

exist for EV chargers. Until such a code is implemented, any EV charging curtailment has to happen on a voluntary basis. If a DSO offers congestion management program through charging curtailment, households have a choice to sign up, and even if they are signed up, infractions are likely to occur. Automatic charging curtailment suffers from the free-rider problem: in a voluntary organisation, where no player can be excluded from the benefits of congestion management, individual players have no reason to cooperate with the group [30]. Therefore it is likely that a DSO has to provide some sort of incentive for participation, and create a sanctioning system for infractions. The size of this incentive is analysed.

As discussed in subsection II-I, Ostrom [30] suggest that sanctions should be graduated and start small. Thanks to the proliferation of smart meters, monitoring of player behaviour (if consent is provided) is possible. It is important to know which player caused the grid failure, and if any infractions were involved. All players are harmed by grid failure, but the offending player should incur additional cost. If information about the offence is accessible to all, then the social cost of the infraction can be a significant deterrent. However, according to design principle 5), it is important that enforcement is applied by an official that is accountable to the players themselves. The perceived fairness of the system and the effectiveness of monitoring and enforcement is an important factor in maintaining quasi-voluntary compliance.

Each of 5 approaches above is evaluated in a realistic grid simulation using reinforcement learning. Each simulation results in an equilibrium, with each player having a set of beliefs about the action with the highest reward given certain information. The next section discusses the grid simulation approach.

B. Grid simulation

Mulenga, Bollen, and Etherden [17] highlights three approaches from literature for estimating the hosting capacity of distribution grids: deterministic, stochastic, or time-series, which respectively require inputs to be fixed values, probability distribution functions, or time series data. A combination of different methods is also possible. However, no mention is made of game theoretic elements in which player demand changes based on the outcome of the game. To understand what the hosting capacity is of a grid in which players are actively maximizing their payoff, a different type of simulation is required, which includes player strategy. Key to the setup of this type of simulation is understanding how players value the outcome of the game, which is also referred to as the 'utility' of the outcomes.

The goal of this thesis is to analyse the influence of various strategies for household power consumption. The outcomes measured are the frequency of current violations, and the level of inconvenience caused to users, as measured by energy not served.

Note that it is not the goal to create a virtual grid that accurately represents a real residential grid. Instead, it is to create a model that is realistic enough to provide plausible incentives for user behaviour. The simulation is focused on maximizing the number of learning opportunities for players, to rapidly converge towards equilibrium beliefs, rather than accurately simulate grid use throughout the day. The focus lies on creating realistic dynamics for rapid learning, rather than a grid model that has full external validity. Only weekdays are simulated.

1) Simulation environment: To achieve this goal, a simulation is setup using various open-source tools available for Python. The neighborhood model is derived from the setup conceived by Nijenhuis [57], with some key differences. A commonly occurring feeder (cluster 6) is selected from [57]: an aluminium cable with a cross-section area of 50 mm^2, 290 meters long, connecting 26 households (illustrated in Figure 6. This cluster represents a commonly occurring suburban grid (representing 3.4% of LV feeders) with relatively undersized infrastructure for today's standards. The feeder is modeled in pandapower using the built-in *NAVY 4x50 SE* cable type, which supports a maximum current of 142 A (equivalent to about 100 kW of power). Table III shows further technical characteristics of this cable type. The simulation models all users as drawing from a single phase (or equivalently, three perfectly balanced phases).

In practice, multiple feeders connect to a transformer, and the transformer is undersized with respect to the sum of capacity of the feeders [57]. To account for the transformer limitations, a different simulation could be setup where all households connected to a transformer jointly manage congestion, or a lower current limit for the feeder could be used to imply a transformer that is nearing capacity.

For each interval, the resulting voltages and currents for each node in the simulation is analysed using a rudimentary load flow calculation provided by the open source pandapower software which was developed by Thurner, Scheidler, Schäfer, *et al.* [58].

2) Load profiles: The action set of the players only includes the EV charging decisions. However, charging is not the only load on the grid, as the load is added to a fixed base load for each household. The base load represents the normal use of devices throughout the day, and is relatively low, although short peaks of a couple of kilowatts can occur for cooking or vacuuming.

The Artificial Load Profile Generator (ALPG) was created by Hoogsteen [6] to create realistic load profiles for testing a decentralised energy management approach. The profiles are

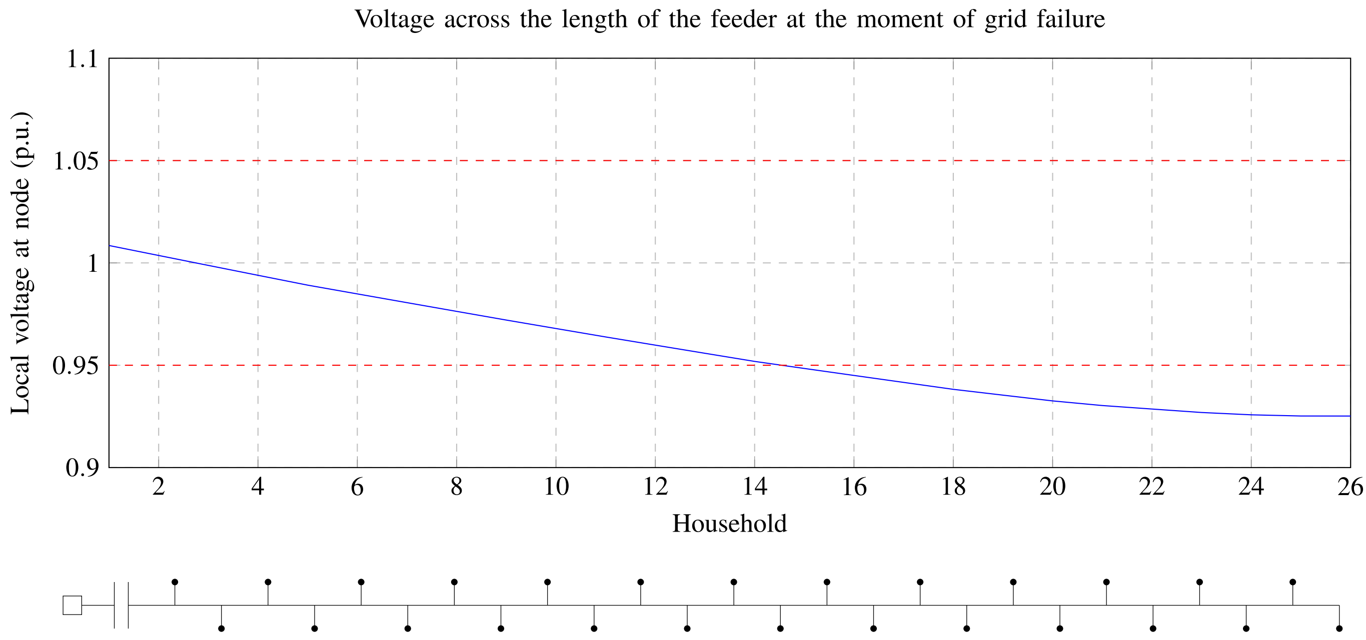


Figure 6: **Bottom**: visual representation of the local power grid with 26 households. **Top**: simulated voltage at the household nodes at the moment of grid failure.

created by simulating the appliance use of individuals living in a household. In this thesis, the ALPG is used for generating a large number of daily baseload profiles.

The ALPG can generate profiles for different types of households; workers, retired individuals, or families. For every household there exists a single adult and double adult variant (see Figure 7 for average smoothed profiles). In this thesis, the two types of households that are considered are the single and double worker households, making up respectively 39% and 61% of the players, which is representative of the Dutch population [59]. 200 weekdays are simulated in the ALPG, 100 days for each profile. These profiles form a pool of options for the grid simulation to randomly sample from.

On top of the baseload profile, the ALPG allows for the generation of start-times of a dishwasher and a washing machine. These are relatively high power compared to the baseload, but low compared to EV charging demand. Practically, the inclusion of these devices increases the level of noise in household demand and the likelihood of current violations.

A core tenet of congestion management is the simultaneity factor (SF) discussed in subsection II-B. To ensure that the the ALPG profiles create realistic levels of congestion, they must adhere to theory provided by Rusck [14]. This can be tested by comparing the actual maximum power within a time window with the expected maximum power $P_{max,N}$ in that time window given by Equation 3:

$$error = \frac{max\left(\sum_{i \in N} \mathbf{v}_i\right)}{P_{max,N}} - 1. \tag{8}$$

If done using the sum of all profiles ($N = 200$), for 24 time-windows of 1 hour, the mean error of the sample is $\mu_s = 0.14$ with a standard deviation of $\sigma_s = 0.16$. That means the peak simultaneous power consumption of households generated by the ALPG is on average 14% higher than expected using the equations from Rusck. In one case, it is 58% higher. This indicates that the equations from Rusck are not very reliable in this scenario, likely because since 1956 household behaviour has changed significantly, with households using more power for longer amounts of time, thereby skewing the relationship between the mean and peak loads.

Nonetheless, the theory behind the SF holds, as the data analysis shows an average simultaneity factor of 0.52, indicating that household power demand is desynchronized to a large extent.

3) EV initialization: Each scenario is evaluated with 100% EV penetration. All the EVs are identical, with a 72.5 kWh battery, and charging power up to 11 kW. Players can only choose between waiting or charging at max power. Nuanced dynamics of EV charging related to efficiency and temperature are not relevant for this research and therefore not included.

The ALPG tool allows for the generation of realistic EV charging sessions for the same household profiles that are discussed previously. For this research, 365 days of EV charging sessions were generated for both the single and double profile, which include an arrival time and an associated charge demand. Weekend days are filtered out. In about one-third of the days, despite it being a workday, no charging takes place. Histograms describing the generated data are displayed in Figure 8. It can be seen that most arrival times are around

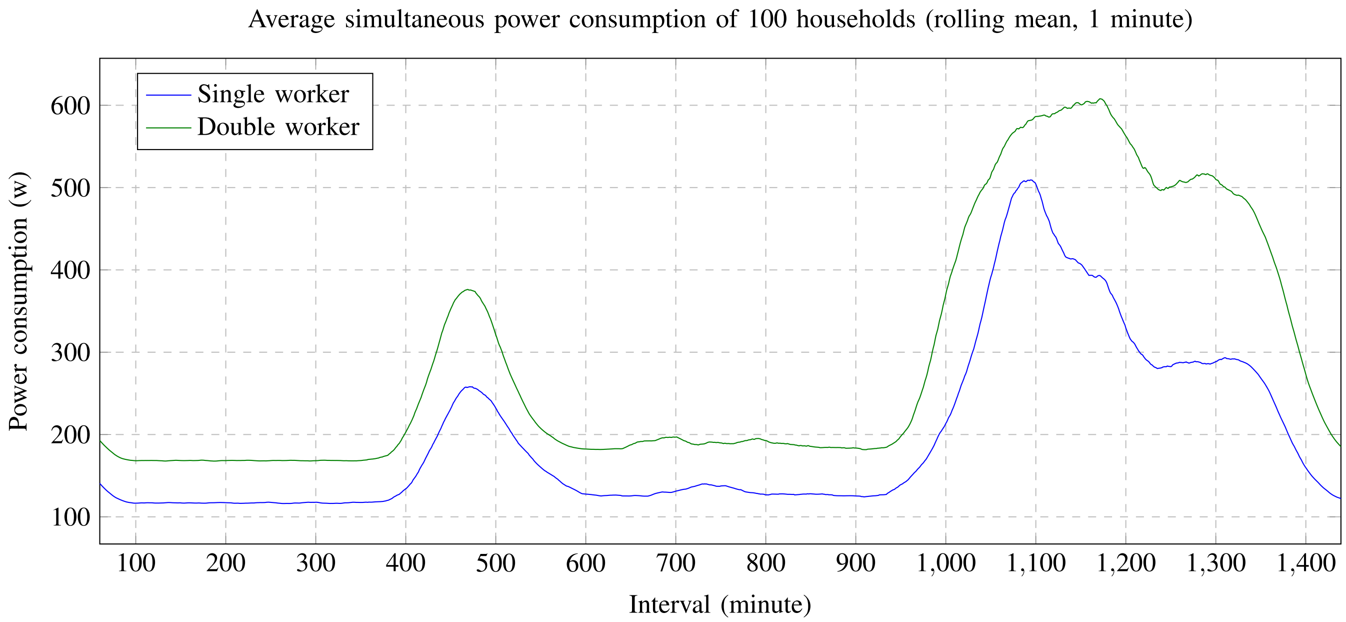


Figure 7: Graph comparing the power profile from different household types in the ALPG generator

interval 1050, which is 17:30 o'clock, with a secondary peak around interval 1200, at 20:00 o'clock. In all cases cars arrive after interval 1000, which is the point at which the simulation starts. The charging demand is most often somewhere between 5 to 20 kWh, but in rare cases can reach up to 60 kWh. The average charging demand is around 12 kWh. The simulation stops at midnight because the goal is to train players for decision-making in the evening, during which new cars can arrive and other high-power appliances are in use.

C. Reinforcement learning

One method highlighted by Bowling and Veloso [54] for finding an equilibrium in a stochastic game is called Q-learning. Q-learning is a model-free reinforcement learning technique introduced by Watkins and Dayan [60]. Model-free means that the learning method does not aim to create a model of the environment, but learns solely from the rewards it receives from taking certain actions given information about the state of the environment. The basis for the Q-learning approach is the individual player's Q-table. The Q-table tracks the expected value $Q(s,a)$ for each available action a in each possible state s. For a given policy π the values in the Q-table are defined using the Bellman equation:

$$Q^{\pi}(s_t,a_t) = r_t(a) + \gamma V^{\pi}(s_{t+1}), \tag{9}$$

where $V^{\pi}(s_{t+1})$ is the expected reward for the next state, discounted by a factor γ. The idea behind this equation is that the available actions should not only be evaluated for their immediate reward, but also for their influence in determining the next state. In the case of the grid congestion game, the players thus receive a reward equal to the energy they draw in the interval after taking their action (about 180 Wh per minute if they are charging, and a couple of Wh per minute for the baseload), plus the reward associated with the optimal action in the new state. If the new state is one that is more likely to transition to a state of grid failure, then it likely has a lower maximum reward than a state where grid failure is unlikely.

The unit of the Q-value can be expressed as the discounted kWh drawn from the grid as a result from the taken action. It contains both the direct reward for the action as well as the discounted reward from the best action in the new state. The results section will elaborate further on the interpretation of the Q-values.

Equation 9 is slightly different from the equation proposed by Watkins and Dayan [60] since in the proposed game the future state is only partly determined by the action of the each player, and affected by stochastic variables such as the EV arrival interval and the baseload of each player.

Equation 9 can further be further modified to include rewards in intervals past the immediate next stage. In the presently considered grid congestion game, players are programmed to not all make a decision in the same interval. Rather, there is a given chance p_c that a player makes a decision during an interval, resulting in an average of $1/p_c$ intervals between decisions. If the grid fails in this period, even if this is not right after the decision was made, the player should take into account the contribution of their action to the failure of the grid, subject to the discount rate. The modified Q-value definition includes the sum of discounted rewards up until a new decision is made by the player (after i_{new} intervals):

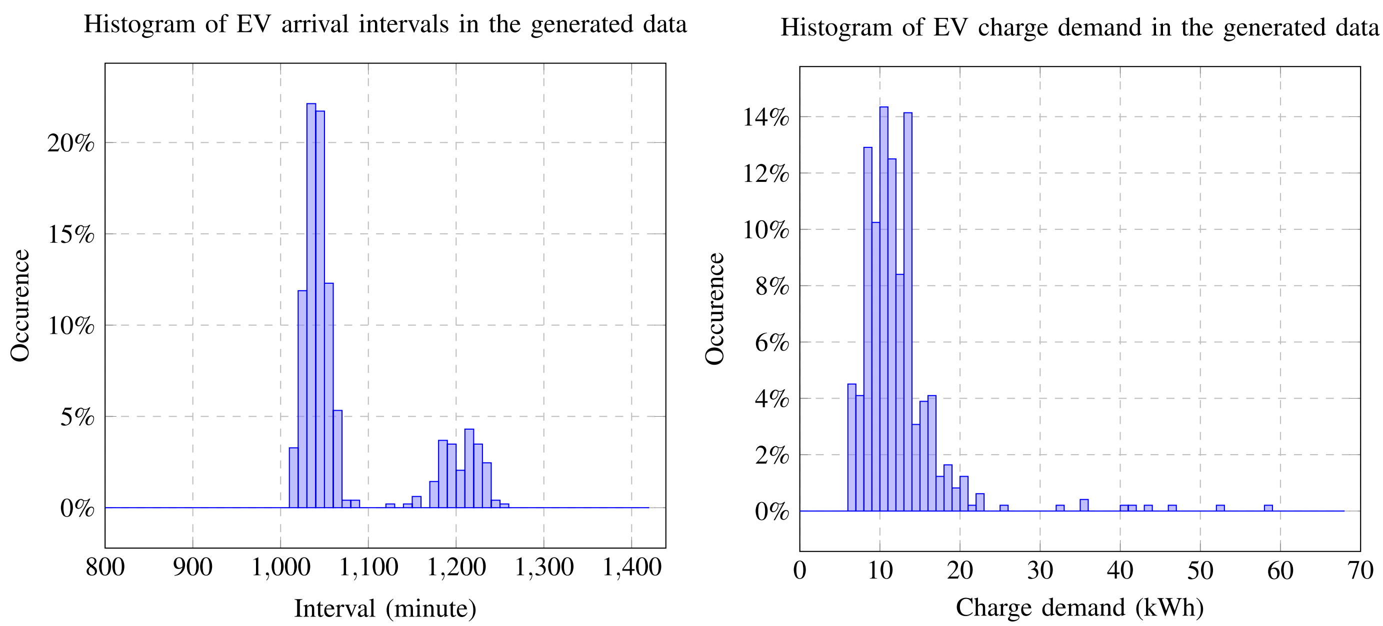


Figure 8: Two histograms showing the distribution of arrival intervals and charge demand in the EV data generated by the ALPG

$$Q^{\pi}(s_t, a_t) = \left[\sum_{i=0}^{i_{new}} \gamma^i \cdot r_{t+i}\right] + \gamma V^{\pi}(s_{t+1}). \tag{10}$$

The Q-learning process is performed using a Monte-Carlo simulation, which, after allowing each player to make a high number of decisions and experience the corresponding action-reward combinations, is used to converge to a (local) equilibrium of each player's Q-table (and equivalently, each player's beliefs). Players take different actions based on the given 'exploration rate' ε. For each action taken, each player updates their Q-table, according to the following function:

$$Q_{new}(s,a) = (1-\alpha)Q_{old}(s,a) + \alpha\left(\left[\sum_{i=0}^{i_{new}} \gamma^i \cdot r_{t+i}\right] + \gamma \cdot \max_a Q(s_{t+1}, a)\right), \tag{11}$$

where α represents the learning rate. The implemented Q-learning algorithm is displayed in Algorithm 1. This algorithm allows for the gradual convergence towards an equilibrium over a large number of iterations. It should be reiterated that this approach does not aim to accurately model real grid user behaviour but instead aims to determine the equilibrium strategies of players in a semi-realistic setting.

In the 'GridShield' scenario, the algorithm is adjusted in that players have 3 actions: *wait*, *charge*, and *charge with GridShield enabled*. If, during the determination of the new grid state, a current violation occurs, all players with GridShield enabled uniformly reduce their charging demand, either until the current violation is resolved or the players cannot reduce their demand any further.

Algorithm 1 Pseudocode for converging Q-values through Monte-Carlo simulation

```
Initialize grid
Assign household size and EV ownership
Randomize Q-tables
for repeat ← 1 to number of repeats do
    for day ← 1 to number of days do
        Reset the grid
        Sample player baseload
        Sample EV arrival time and EV demand
        for interval ← 1000 to 1439 do
            Determine grid state and ε
            Get player power demand        ▷ see Algorithm 2
            Determine new grid state and α
            Determine rewards and punishments
            Update player beliefs          ▷ see Equation 11
        end for
    end for
    Save repeat data                       ▷ for analysis
end for
```

1) Parameter selection: The value for the discount rate γ is an important determinant for the outcomes of the simulation. For a low value of γ, players are more motivated by the immediate reward, and less by the reward that could be obtained in the future. The discount rate can be interpreted as the level of uncertainty players have about the future. If a game has a likelihood of p_e to end in any given stage, then a reward-maximizing player will multiply the reward they could expect in that future stage game by the probability of reaching that state. This probability decays exponentially as players predict further into the future.

Ostrom [32] claims that a low discount rate is is key to the sustainable management of CPRs. A low discount rate implies that users give high value to rewards that are far into the future, which incentivices strategies that provide long-term stability of the resource. A low discount rate would indicate that players don't believe in the value of long-term stability, because (1) players believe that the resource is doomed to exhaustion regardless of their actions, or (2) they are not interested in long-term exploitation of the resource. In a stable electricity grid, the likelihood of failure is low and the discount rate can be expected to be very high, and is modeled as such in this thesis ($\gamma = 0.95$). In reality, this simplification could be problematic, as grid failure and player behaviour form a **negative feedback loop**. If grid failure occurs, players might lower their internal discount rate to account for this, resulting in short-term strategies that focus on charging as much as soon as possible, causing grid failure to occur even earlier. This feedback loop would prevent convergence towards behavioural equilibria where the grid is stable.

Two more variables in this algorithm need to be discussed, the exploration rate ε and the learning rate α. The exploration ε rate determines how often players will deviate from what they believe is the 'optimal' action, e.g. by charging even though they believe waiting would yield a higher reward (see Algorithm 2 for the implementation of this logic). Exploration is very important for determining Q-values, especially at the start of the simulation, when the initial beliefs are randomly initialized. One can imagine the explorations as (1) inexperienced players actively testing actions to acquire information about the environment or (2) players breaking the rules because of unique external circumstances that temporarily impact their internal evaluation of the reward. Some level of exploration is therefore expected at all times. However, too much exploration would harm the convergence towards a behavioural equilibrium. As all players are creating policies contingent on the actions of others, some level of action stability is required to allow Q-values to converge.

The learning rate determines the amount the Q-value is updated after each action. A high learning rate leads to faster convergence, but can also cause Q-values to overshoot the equilibrium value, and can give high influence to 'noisy' results. A lower learning rate allows for the combination of many outcomes over time.

Both α and ε need to be set to a value where they allow for the convergence to an equilibrium as fast as possible, while being low enough to provide stability for the learning process of the individual players. An exponentially decaying function allows for the trade-off to be fine-tuned for a good trade-off between fast convergence at the start of the simulation, and more nuanced evaluation of rewards later in the simulation. The chosen functions for α and ε are dependent on the number of decisions the player has made so far, and are shown in Figure 9. For experimental setups where the Q-values are not randomly initialized, such as the (4) rotation system, both α and ε is fixed at a low value.

The chance of making a decision in a given interval p_c for 26 players was set to $p_c = 1/(2 \cdot N) = 1/52$. On a single simulated day, players are expected to make around 7 decisions each. The chance that a number of players makes a decision simultaneously can be calculated using a binomial distribution:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}, \tag{12}$$

where n is the number of players, and k is the number of players making a decision. It is obtained that:

$$\begin{aligned} P(X = 0) &= 0.60, \\ P(X = 1) &= 0.31, \\ P(X > 1) &= 0.09. \end{aligned} \tag{13}$$

The chance that more than 1 player makes a decision during an interval is 9%. This causes a considerable amount of noise for players evaluating the reward for their decision.

2) State description: Charbonnier, Morstyn, and McCulloch [61] provide an example of Q-learning applied to a local distribution grid, albeit as an optimization approach, not an approach for evaluating grid user behaviour. They highlight the issues caused by dimensionality and stochasticity for reinforcement learning. As Q-learning is model-free, it needs to learn about the rewards for given actions for every possible state individually. As more state variables are added across additional dimensions, a larger number of simulated decisions is required, increasing convergence time. In the presently considered grid congestion game, users can be informed about grid properties such as load and voltage, which are continuous variables, and suggesting a large number of possible states.

The use of categorical instead of continuous variables can be justified in this context. Firstly, of the many states that the grid can represent, the states that are associated with grid instability are of particular importance for the learning process. During periods of low demand players can freely choose between actions, and don't have to be concerned with their impact on the grid. Secondly, from a behavioural perspective, categorization is justified because users of any system mentally discretize a continuous variable into 7 ± 2 categorical variables, such as 'low', 'high', or 'very high' [62]. The state information provided to players in different congestion management approaches is summarized into a few categorical states to reduce training time and to promote the clarity of outcomes.

D. Hypotheses

A methodology is proposed in which a reinforcement learning technique is used to evaluate the incentives that are visible to players sharing a local distribution grid. Players represent households, whose power draw and EV demand is determined by randomly sampling from a pool. At random

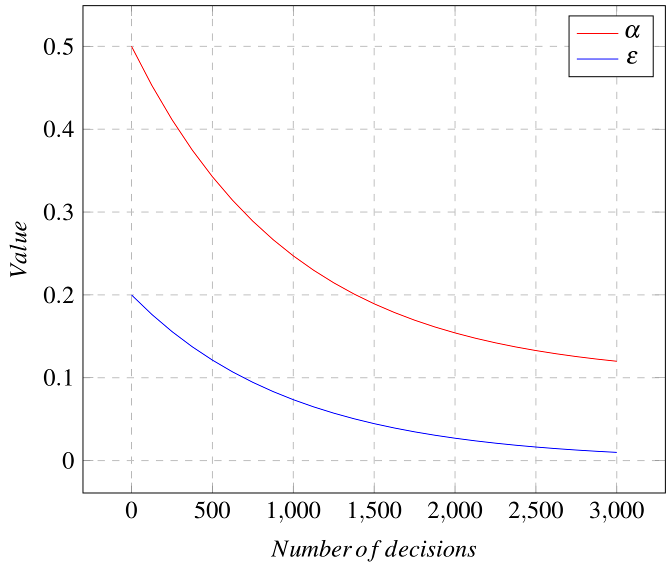


Figure 9: α and ε decrease over time

Algorithm 2 Pseudocode describing player logic for choosing an action

```
for player in players do
    Generate a random number r_1 between 0 and 1
    if player has an EV available for charging
      and r_1 < action probability p_c then
        Generate a random number r_2 between 0 and 1
        if r_2 < ε then
            Choose a random action a
        else
            Select a = argmax_{a'} Q(s, a')
        end if
    end if
    if charge in a then
        Add charging demand to energy demand
    end if
end for
```

intervals, players choose an action based on the information available to them. The reward for that state-action combination is evaluated using the Q-learning algorithm. Grid simulation is used throughout for evaluating the grid state.

This technique can be used to answer questions about the design of a cooperative approach to congestion management. Three hypotheses are tested:

- H1: Rationally experimenting players of a grid congestion game provided with up-to-date information on the grid can learn to prevent congestion;
- H2: Any ruleset that does not plan for near-maximum exploitation of the resource, creates incentives for deviation;
- H3: Incentives and/or sanctions are required to prevent free-riding behaviour in coordinated grid congestion management approaches.

Hypothesis H1 tests whether the incentives in the proposed game actually allow for congestion management by players. It is possible that a 'tragedy of the commons' situation occurs where self-interested players aren't able to achieve a mutually beneficial outcome. Different types of grid information can be provided to players, such as the grid load or grid voltage.

Hypothesis H2 follows from logical reasoning about congestion management. Any congestion management approach in a saturated grid requires the reduction or shifting of power demand, but players still aim to maximize their reward by using as much power as is available without causing grid failure. The expectation is then that in a behavioural equilibrium, players will try to collectively push the grid load near the maximum, while preventing exceeding it. If a significant portion of grid capacity is unused, then incentives exist for players to change their behaviour.

Hypothesis H3 regards the necessity of sanctions and arises from the theory on commons management. In a situation where players have created a ruleset in which they limit their appropriation from the shared resource, opportunities for free-riding exist. According to Ostrom's design principles, a combination of monitoring and sanctions is required to prevent this behaviour [30]. The sanctions should be gradual, but should at least nullify the additional reward a player obtains from an infraction.

These hypotheses will be evaluated through the four scenarios that are tested using the proposed methodology.

IV. Results

The reference point for the EV hosting capacity of the considered grid layout with 26 households is given by Nijenhuis [57], who found that this grid, in an (1) uncontrolled condition, supports a maximum of 5 11 kW chargers or 16 3.7 kW chargers.

A. Load informed

The congestion management approaches based on providing players with basic information about the state of their local grid have been evaluated primarily to confirm that the proposed method yields understandable and logically sound results, and in the second place to confirm hypothesis H1. Approach (2) provides players with information about the grid voltage at their connection point and approach (3) provides players with information about the grid load at the transformer. The approaches are competitive in nature, in that all players are provided with the same information, and no collaborative policies are possible.

The graphs that are most straightforward to interpret are those from the load informed approach, where the players are provided with 4 states representing that the load of the grid measured at the transformer is below 100%, 90%, 70%, or in case of grid failure, 200% of the maximum load. Figure 10 shows a heatmap of the resulting values after running the convergence simulation of Q-values 100 times. Table IV shows the average propensity to charge for all players at each state. Figure 11 shows the progression of all runs over the course of their individual simulation, overlaid in one graph, and can be used to determine whether the resulting equilibrium is deterministic. Lastly, Figure 12 shows a heatmap of when grid failures occurred most often over the course of the training.

The meaning of the final expected rewards (the Q-values) for each player is not immediately apparent. A part of the expected reward is given by the actual amount of energy that a player appropriates from the grid, discounted by γ over time. We can calculate the discounted reward for a fixed load for a long period of time (approaching) using the sum of an infinite geometric series:

$$\sum_{k=0}^{\infty} r\gamma^k = \frac{r}{1-\gamma}. \tag{14}$$

For example, for a charging car the reward is around 180 Wh per minute. Using a discount value of $\gamma = 0.95$, the multiplier for the infinite geometric series to is equal to 20, giving us a discounted value for charging to infinity of 3,600 Wh. However, this amount is not even near the Q-values that are shown. Next to the direct reward, the Q-values are made up of the discounted maximum return in the new state. Since the states all refer to each other or themselves, the Q-values need to again be multiplied by 20, giving us a maximum value of 72,000. Since all actions don't fully charge to infinity, but stop earlier or are affected by grid failure, the final values are lower, in the range of 10,000 to 40,000.

The difference between the Q-values for waiting and charging is given by both the additional reward for charging, and the reduction in reward which is caused by the higher chance of grid failure when charging. Punishment, if present, also affects the difference between Q-values.

It can be seen in Figure 10 that the Q-learning approach has been able to distill some obvious facts. When the grid is in a failed state, generally players will choose to wait to contribute to restoration of power. When the grid load is between 90% and 100% of maximum capacity, the addition of a charging car (which draws 11 kW) will almost always lead to grid failure (unless another player stops charging at the exact same time), and thus players have generally realized that they should not charge in this state. When the grid load is below 70%, charging is a safe decision, unless 3 players decide simultaneously to charge (a 1% chance).

Figure 12 shows that players are successful in reducing the amount of grid failure in the start of the evening, but that this task becomes more difficult once the second wave of EVs arrives, around interval 1200. This is an indication of the non-linear impact of EV penetration on grid failure. Once a given threshold of EVs within a single feeder is crossed, congestion becomes more difficult to manage.

In the progress overview in Figure 11, it can be seen that, although failure decreases dramatically at the start of the simulation, it starts to build up towards the end. To confirm the long-term convergence of the simulation, 2 runs have been extended to 60 repeats. It can be seen that although failure stabilizes at a higher point than the minimum, which occurs around repeat 5, it contributes to a lower overall unserved EV demand. A possible explanation for this comes from the dynamic nature of the game. Initially, players learn to back off during periods of high load, to avoid grid failure. As the grid becomes more stable, players learn that, occasionally, charging during high loads can be beneficial. Surprisingly, as shown in Table IV, 22% of players learn to expect higher rewards when choosing charging when the grid is in a failed state. It is possible that players learn to expect that other players, making a decision at a later point, will reduce their demand during grid failure, as to enable rewards from power draw.

Increasing one's load during grid failure is a type of counter-intuitive, free-riding behaviour of players that damages the reliability of the socio-technical system. However, the emergence of a sub-group of players which exhibit this type of behaviour is consistent with the *hawk-dove model* in evolutionary game theory [63]. This model describes the interaction between selfish (hawk) and prosocial (dove) behaviour in a game with a limited amount of resources. It is shown that, depending on the parameters of the game, an equilibrium is reached with a fixed ratio of both types of strategies. Although players behaving aggressively obtain a higher reward, they are reliant on the existence of a large group of prosocial players.

The results show that the load-informed congestion management scenario creates incentives for the emergence of a (small) group of selfishly acting players. Could this type of behaviour occur between real households? It depends on the adopted ruleset; if resolving grid failure is really as simple as the reduction in load of a few players, then yes, it is likely that some players will start to expect others to resolve grid issues, without themselves contributing to a solution. In practice, the 'punishment' of grid failure will likely be perceived as more severe, considering the impact it has on a household's functioning, which will disincentivise any policies that rely on grid failure to occur.

From these results, hypothesis H1 can be confirmed, although some limitations need to be kept in mind. Q-learning with grid load data indeed allows the players to manage congestion as to minimize unmet demand. However, due to the structure of the games and its incentives, minimizing the unmet demand allows for a non-zero level of grid failure.

B. Voltage informed

As previously seen, players provided with grid load information can do a decent job at stabilizing the grid, going from 160 to 25 minutes of grid failure per day. When provided only with information of the voltage at their local connection to the feeder, players do not fare so well.

Figure 13 shows the heatmap of resulting Q-values after 58 runs, for 4 states representing the voltage under 1 p.u., 0.98 p.u., 0.96 p.u. and 0 representing grid failure. It was discussed that local voltage is a bad proxy for grid load, as the voltage sag is more pronounced towards the end of the

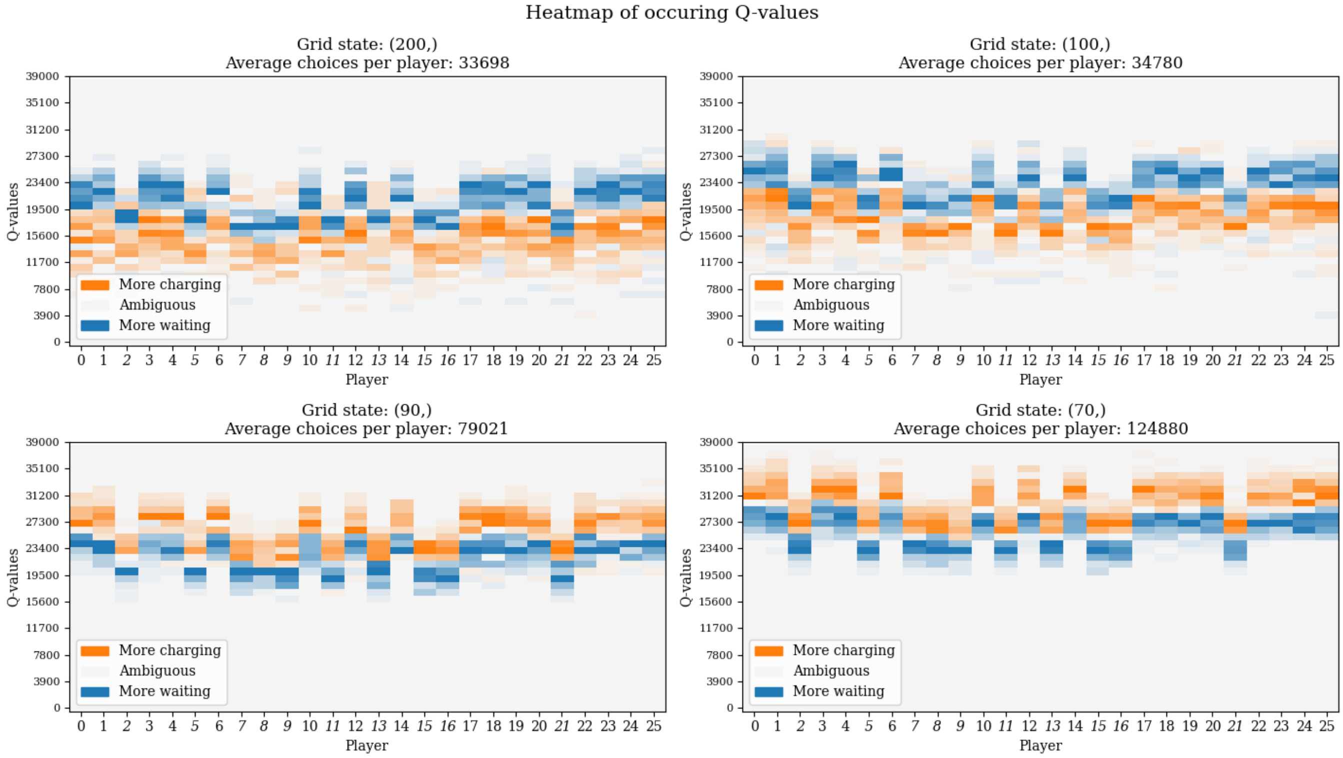


Figure 10: Load informed: heatmap of final Q-table values for each player after 100 simulations

Grid state	Mean	Standard deviation	Average choices per player
(200,)	0.210000	0.407308	33698
(100,)	0.210000	0.407308	34780
(90,)	0.860000	0.346987	79021
(70,)	0.990000	0.099499	124880

Table IV: Load informed: mean preference to charge of all players for each grid state after 100 simulations

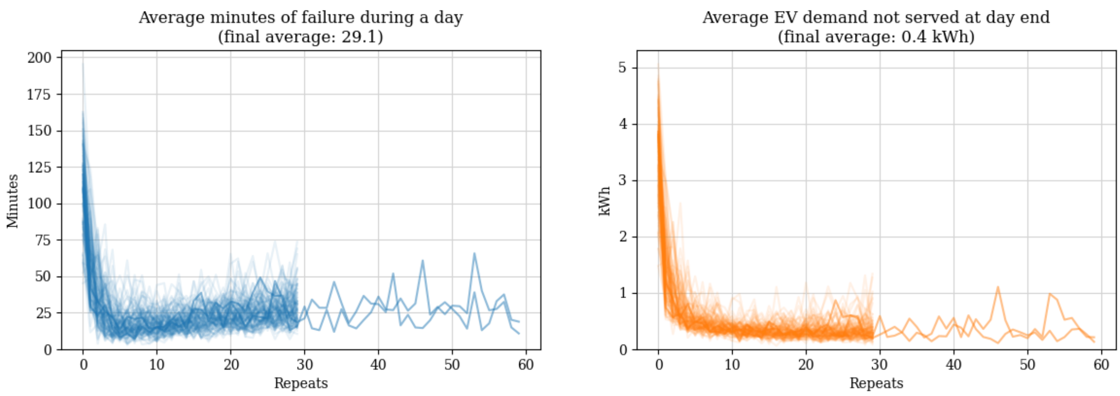


Figure 11: Load informed: summarized progression of all 100 runs

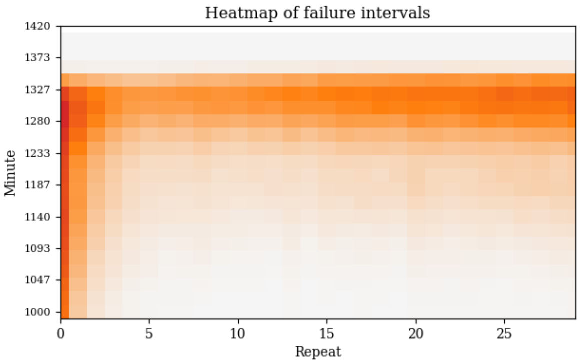


Figure 12: Load informed: heatmap of failure intervals

feeder. This can also be seen in the heatmap; at lower voltages, the heatmaps show more convergence at the end of the feeder, while at the start of the feeder the Q-values haven't moved much from their initial random initialization. This is because these lower voltages do not occur close to the transformer. Without experiencing the voltages, players can't learn about the impact of their actions.

According to Table V, players are unanimous about choosing to charge when the voltage is above 0.98 p.u. (corresponding to state 1.0). For the lower voltages, players are more divided about the best course of action, and show higher standard deviations. Similar to the results for the load informed approach, 31% of players learn to expect higher rewards for choosing charging during a grid failure (state 0.0).

Figure 14 shows the progression made by players over time. Although players are able to reduce the amount of failure and unserved EV demand initially, further training does not improve outcomes significantly. Figure 15 shows that players are able to reduce the amount of grid failure at the start of the day significantly, but that once the second wave of arrivals starts around interval 1200, grid failure cannot be averted.

C. Rotation

In the rotation approach the evening is divided into half-hour slots, which are assigned to randomly created groups of households. The number of groups was varied to evaluate the effect on game outcomes. Players are provided with limited information about the state of the grid (only a 'high' and a 'normal' state), as well as information about whether it is their turn to charge, giving us 4 unique states.

Unlike the previous two approaches, the rotation approach requires coordinated action from all players. Only if (almost) all players stick to the assigned schedule, are the benefits of the scheduling visible. Players cannot learn to stick to the schedule if they start with randomized Q-values. Therefore, the Q-values will be initialized to reflect the belief that sticking to the schedule yields higher benefits. Additionally, the high level of exploration used to speed up the convergence in previous approaches would cause rapidly fluctuating behaviour and loss of belief. Therefore, α and ε are both initialized and fixed at 0.1. The simulation can show if the incentive structure allows players to maintain their beliefs about the best course of action.

Next to the number of groups, the other variable that was tested for in this approach is the size of the punishment that is applied to players that charge when it is not their assigned turn. This punishment is subtracted directly from the Q-value, and is therefore expressed in the same units (discounted Wh). The practical meaning of the punishment will be discussed in subsection V-B.

Each combination of group size and punishment was simulated for 20 repeats of 20 days. The average minutes of grid failure and unserved EV kWh per day of the last repeat are displayed in Figure 16. Because only a single simulation was run per scenario, the resulting lines are relatively noisy. Nonetheless, meaningful conclusions can be drawn.

For low levels of punishment (< 100), none of the given group sizes is able to eliminate grid failure, meaning that players end up breaking the rules in every scenario. In both graphs, a turning point is visible around a punishment size of 180, after which players learn to wait for their turn. As discussed, the difference between the Q-values for waiting and charging is affected by the additional discounted power draw for charging until the next decision and the likelihood of grid failure for each decision. It seems that the a punishment around 180 Wh is adequate to stabilize player beliefs.

However, although low punishments are associated with higher levels of grid failure, unserved EV demand is relatively low. As punishment increases, players are less keen on charging, causing unserved EV demand to increase. Note that players optimize for overall power draw, not just EV charging.

The minimum number of groups required to get grid stability using this approach is 4. This means between 6 and 7 players are allowed to charge at the same time, respectively drawing 66 or 77 kW. Two interesting effects are visible for the number of groups. The scenarios with a larger number of groups show more stability both with low and high levels of punishment. Secondly, a larger number of groups leads to a higher unserved EV demand, as fewer players are allowed at the same time.

Hypothesis H2, which states that near-maximum exploitation is required to incentivise cooperation, can be rejected, as it can be seen that the increase in number of groups above 5, which lowers exploitation, does not lead to worse grid failure outcomes.

An optimal combination can be found that stabilizes the grid while minimizing the unserved EV kWhs. A punishment around 190 with 4 or 5 groups seems appropriate. The Q-values for 4 groups with a punishment of 190 are shown in Figure 17. Comparing Q-values for various group-punishment combinations, it becomes clear that to stabilize the grid, players need to stick to their assigned turns not only when the grid load is high (between 100 and 80), but also when grid load is lower than that (below 80).

The states 100 and 80 were chosen to allow players to

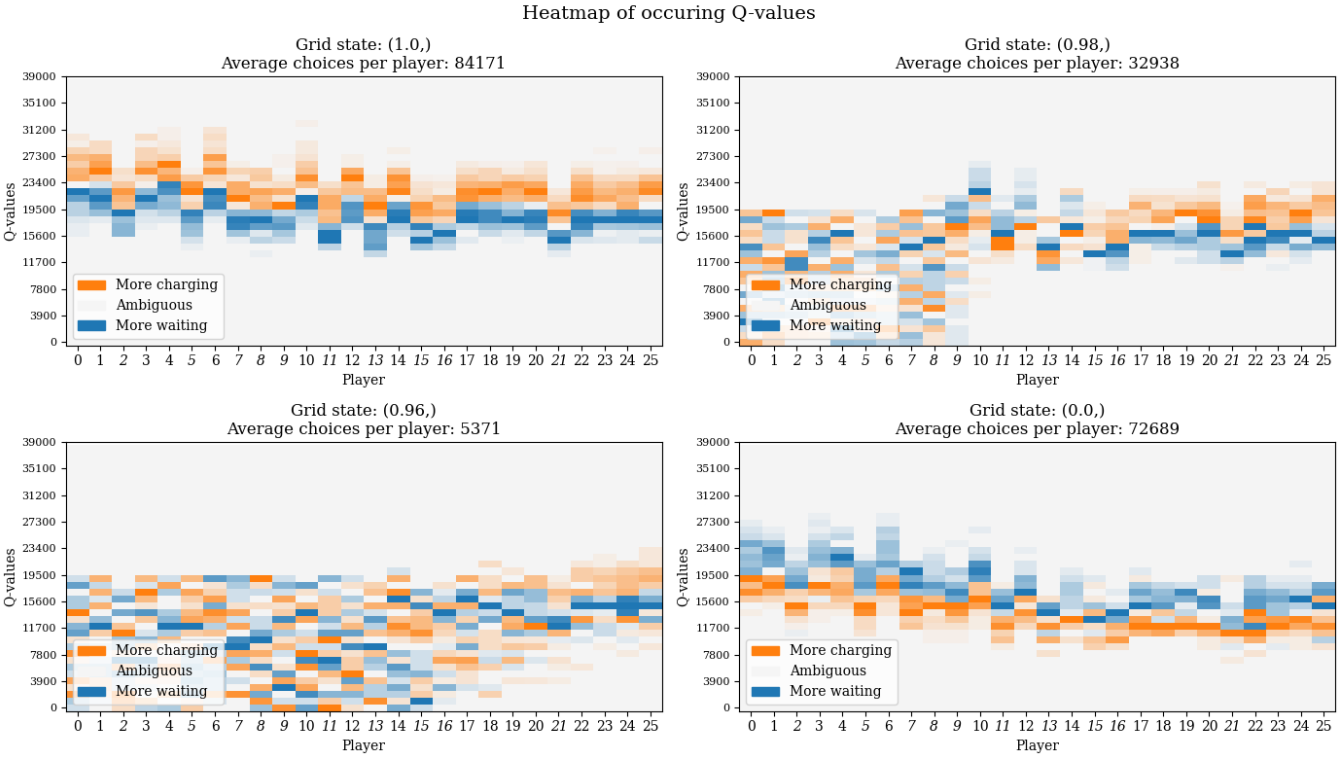


Figure 13: Voltage informed: heatmap of final Q-table values for each player after 100 simulations

Grid state	Mean	Standard deviation	Average choices per player
(1.0,)	0.990000	0.099499	84171
(0.98,)	0.840000	0.366606	32938
(0.96,)	0.550000	0.497494	5371
(0.0,)	0.370000	0.482804	72689

Table V: Voltage informed: mean preference to charge of all players for each grid state after 100 simulations

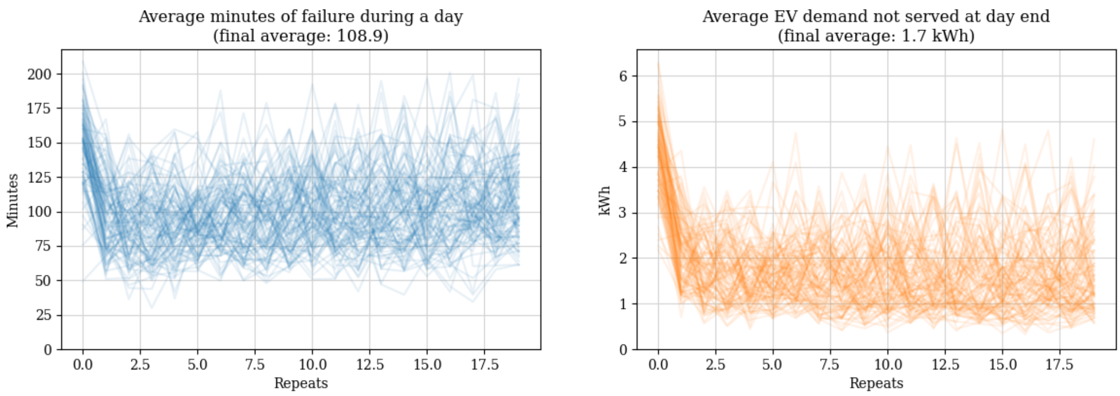


Figure 14: Voltage informed: summarized progression of all 100 runs

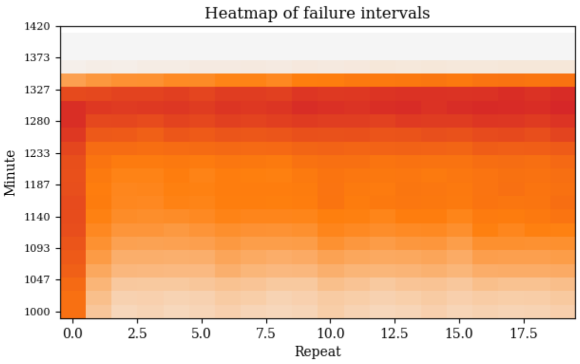


Figure 15: Voltage informed: heatmap of failure intervals

distinguish between periods of high demand and normal demand. Different information (and associated states) could allow players to come to different conclusions, but the general requirement to enforce assigned turns in most intervals would remain. The determined punishment of 190 means that players are expected to 'hand in' part of the reward gained from charging out of turn. This is the most basic form of sanctioning which occurs in CPRs analysed by Ostrom [30], on top of which graduated sanctions should be applied.

The results from this coordinated approach confirm H3, suggesting that some level of incentive is required to enforce a coordinated rotation schedule for EV charging.

D. GridShield

The final approach tested is number (5), using the GridShield load curtailment approach. In this approach, players are given the choice between 3 actions, namely: *wait*, *charge*, and *charge with GridShield*. Players are given information about the grid using 2 states: 'normal', and 'grid failure'.

The expectation in line with H3 is that the GridShield approach enables free-riding behaviour if participation is entirely voluntary. To confirm this 10 runs were done where no incentive was provided. The progression is shown in Figure 18. It is clear that there is no consistent learning progress. In one case, players are suddenly able to coordinate, as can be seen by the line that drops to 0 for both failure and unserved EV demand. Upon closer inspection of the Q-values for this result, as shown in Figure 19, it can be seen that players have settled on a division where 8 players charge fully, while the remaining players use GridShield. This division was settled on entirely randomly during the simulation, but since it eliminated grid failure from the simulation, it was immediately preferred by all players. The outcome is unfair, with 8 players being allowed to charge as much as they want, but outcomes are nonetheless quite good for all players when looking at the unserved EV demand at the end of the day.

The same outcomes could of course be achieved if all players used GridShield, which would fairly distribute the burden of congestion across all players. To achieve this, some sort of incentive can be considered. Various levels of punishment were tested, where the punishment is applied to players choosing to charge without gridshield. The results are shown in Figure 20.

Although the results are mixed at punishments below 35, it is clear that from that point onwards the grid is highly stable, as all users have switched to GridShield. Some level of unserved EV demand remains, but this is due to the demand exceeding what is possible to charge before midnight, and would reduce to 0 if a whole night was simulated.

What is especially notable is that the size of the punishment required is about 5 times smaller than the size of the punishment needed for enforcing the rotation approach. This implies that this system would generally be cheaper, for users in case of a punishment, or for the DSO in case an equivalent reward for participating is instituted.

V. Discussion

The previous section has addressed the simulation results for 4 different cooperative congestion management approaches within a specific neighborhood. The results show that rationally experimenting players, given that they are provided with high quality information about the grid state, can learn to prevent congestion, confirming hypothesis H1. An example of high quality information is the loading of the feeder at the transformer, while voltage measured at the point of connection for the households does not provide adequate information.

Alternatively a ruleset could be implemented without additional infrastructure (rotation), or with additional infrastructure for automatic curtailment (GridShield). The results showed that these rulesets do not have to achieve near-maximum exploitation of the resource to be effective, thereby rejecting hypothesis H2. Grid stability can be achieved if a sufficient incentive exists for adhering to the ruleset, confirming hypothesis H3.

This section dives into the limitations, interpretation, and implications of the research findings. Suggestions are made for future research.

A. Limitations

The methodology proposed in this thesis has been tested on a single test setup, which represents a commonly occurring feeder in the Dutch distribution grid. The resulting outcomes and calculated punishment size will differ per test setup. Even within the tested setup, some assumptions and simplifications have been made.

The first simplification made is that price incentives have not been considered. Although some households in the Netherlands have a contract where the electricity price is fixed throughout the day, others have contracts with peak and off-peak pricing or even variable pricing per hour. Even if users have a contract with fixed prices from their energy retailer, they may have a contract with an EV aggregator that incentivizes them to charge at certain times of the day. Considering that EV demand is a significant share of household energy consumption, price incentives can significantly affect user behaviour.

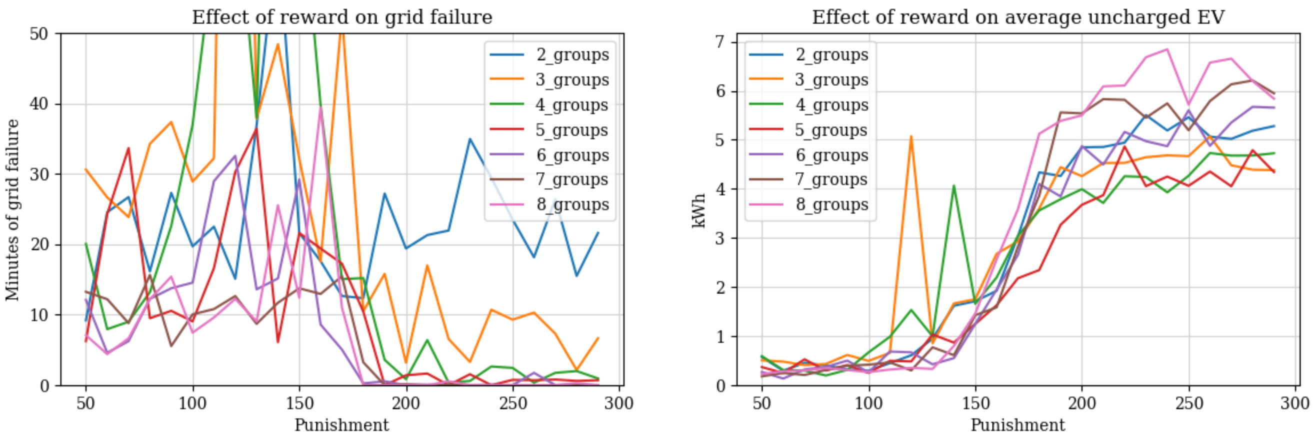


Figure 16: Rotation: Comparing the effect of punishment size accross groupings

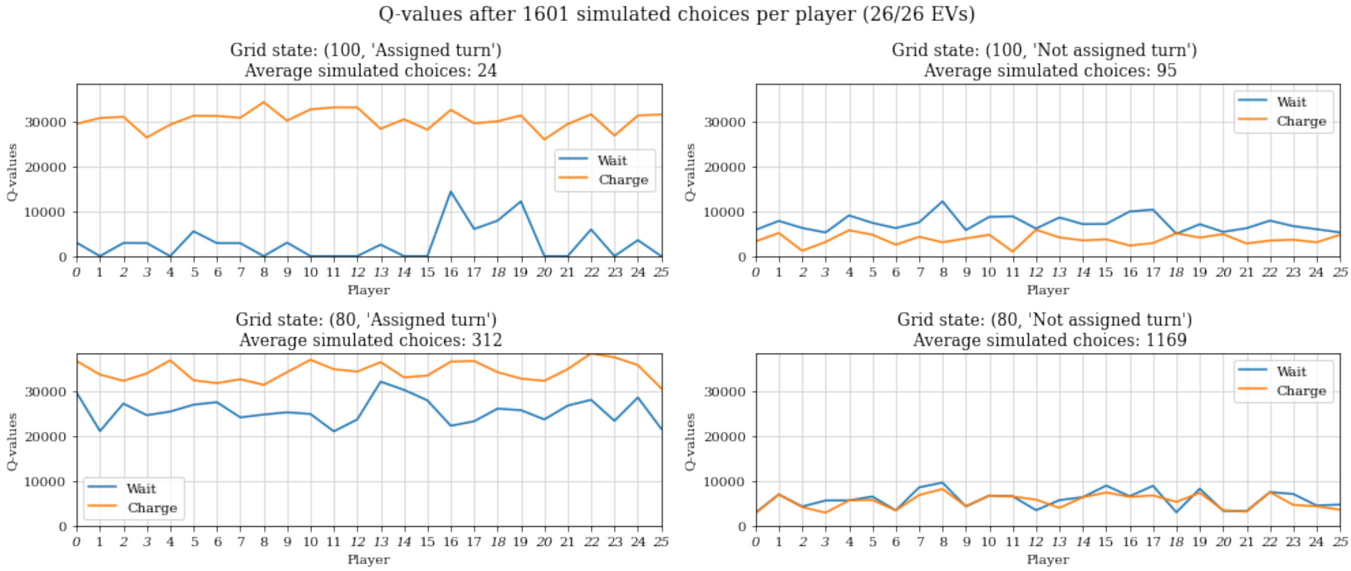


Figure 17: Rotation: Q-values for recommended group-punishment combination

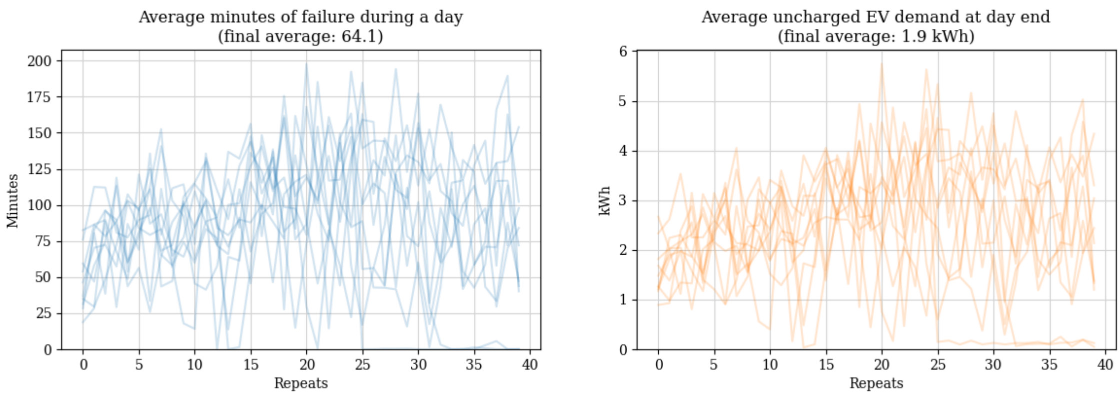


Figure 18: GridShield: 10 runs offering voluntary participation in GridShield

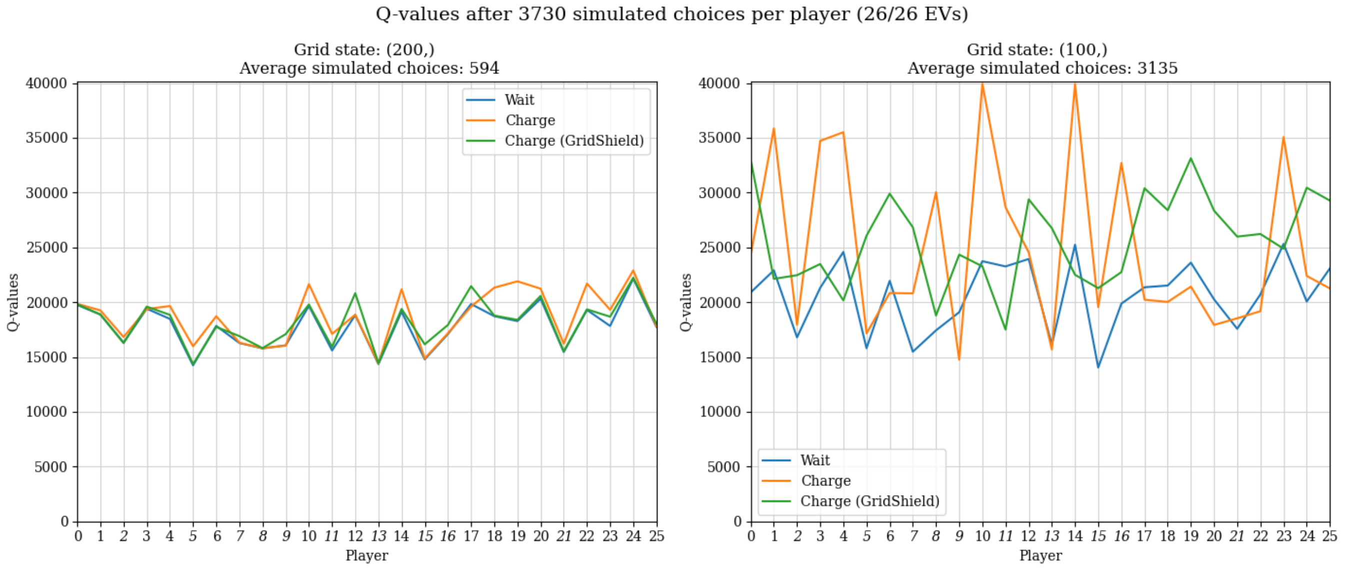


Figure 19: GridShield: inspecting an accidentally succesful run

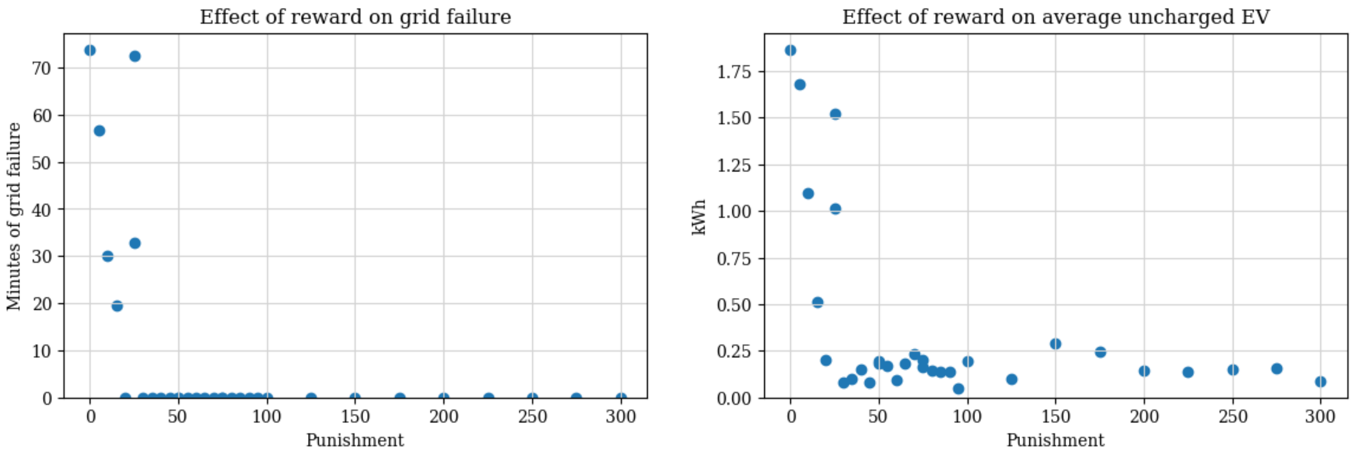


Figure 20: GridShield: showing the effect of punishment size on grid failure and unserved EV demand

One way to address the influence of prices is to adjust the reward structure of the game. Instead of maximizing their power draw, players could focus on minimizing the total cost of their power draw. This would require providing the players with additional information about pricing, expanding the set of states S. To ensure players do charge, remaining EV demand at the end of the simulated period can be counted as a penalty, representing the additional cost of having to charge on-the-road versus charging at home.

Another simplification made in the methodology of this thesis is the limited action set, with exclusively the options to charge or to wait. In real life, households could perform a combination of different actions, including the delayed use of appliances, or charging at a lower power level. This could allow players to more carefully increase their power demand if the grid is nearing capacity. However, additional grid states would be required to allow players to make these fine-tuned decisions.

Both the inclusion of prices and the extension of the action set expand the size of the Q-table. Since the Q-table for each player is 2-dimensional (states and actions), it grows exponentially, rapidly increasing the amount of required training. For more complicated game structures, the model-free Q-learning approach may not be practical. In reality, players would predict the impact of their actions on the electricity grid and on their rewards, indicating that they have constructed a mental model of the grid. A *model-based* learning technique, such as neural nets, could enable continuous rather than discrete inputs, and could outperform Q-learning on training speed. The downside of such a model-based learning technique is the reduced clarity of player reasoning.

Another limitation is found in the hours that have been chosen for simulation. The simulation stops at midnight, while in reality EVs should be able to charge throughout the night. The choice to stop the simulation at midnight had to do with the changing context of decision-making, as the baseload of homes decreased and no new EVs were arriving. In the simulated scenarios, most congestion occurs around 20:00, when the second wave of arrivals takes places. As can be seen in Figure 12 and Figure 15, failure rarely takes place closer to midnight, since most cars with a small charging demand will have completed charging at this time. The downside in stopping simulation at midnight is in the reduced clarity of rewards, also called the 'end effect'. Player rewards are updated as the intervals are simulated, so the cut-off at interval 1439 reduces the reward for any players making a decision close to midnight. This could be resolved by calculating the remaining expected reward beyond midnight and assigning it to player rewards.

A further limitation is found in the assumption of perfectly balanced phases. In the simulation, each household is assumed to draw from each phase equally. In practice, some phase imbalance is likely to occur, as some high-power devices in the base load are single phase only. Phase imbalance would increase the current on one of the phases, meaning the current violation occurs quicker. In this thesis this would create a marginal difference in outcomes.

Finally, the simulation of the grid for each interval created a significant computational burden, and limited the number of scenarios that could be tested. The grid simulation applied in this thesis was required for calculating voltages, but has limited use otherwise. For load information, an approximate estimation of the total load can be found by simply summing the individual loads. Losses and phase imbalance can be accounted for by lowering the grid load limit. Removing the need for grid simulation would allow for more rapid generation of results.

B. Interpreting the size of incentives

In the rotation and GridShield approaches, a level of punishment is estimated which leads to an equilibrium in which players are able to manage congestion effectively. The punishment is expressed in the same unit as the reward, discounted Wh. How do the calculated amounts relate to policy in practice?

In the case of the rotation system, the punishment is almost exactly equal to the size of the additional reward obtained by charging for 1 interval, about 190 Wh. Please note that this similarity is *coincidental*, as the reward includes the sum of many intervals. To administer this punishment, the user could be prevented from charging. This punishment is for a single decision, not for each interval during which the chosen action is active.

The simulated punishment assumes that each infraction is immediately noticed and sanctioned. In practice, this does not have to be the case. As described by Ostrom [30], the sanction should be divided by the chance of enforcement. Depending on the ruleset created by the users, and the level of information shared, different situations can exist. Enforcement could only be limited to cases of grid failure, where an investigation could find the rule-breaker. If enforcement only occurred under limited circumstances, the size of the punishment would have to increase for the same effect on the calculated reward of rule-breaking behaviour. The enforcement frequency could form a significant hurdle for the GridShield implementation, as the defection of a single player does not cause grid failure and would likely go unnoticed, unless stringent monitoring is implemented.

The discount rate γ also needs to be considered. A punishment that consists out of preventing a user from charging would be discounted significantly if it takes place far away from the moment that the player is causing the infraction. The required punishment would grow exponentially, the further away from the infraction it takes place. This calls for a rethinking of the discount rate, as the exponential decay likely does not reflect user's perception of the value of power in the future well.

Considering both the enforcement level and the discount rate, the punishment size that is mathematically correct for the given game rules could reach an extraordinary size, which does not have any practical purpose. Ostrom [30] argues that punishment should rely mostly on loss of reputation, and should be graduated, so that the punishment for a first or second infraction counts more as a warning and a proof that the enforcement system is effective rather than an actual incentive. An example of a graduated punishments would be to remove charging rights from the user that is caught rule-breaking for a number of days, where the number of days increases for repeat offences.

If, as suggested in the previous subsection, the reward structure were based on cost rather than appropriated distribution capacity, then the punishment could also be financial in nature rather than based on appropriated capacity. According to Carlsson, Kataria, Lampi, *et al.* [64], grid users in Sweden in 2017 were willing to pay (WTP) on average 7.5 SEK (€0.80) and 29 SEK (€3) to avoid respectively a 3 min and a 1 hour unplanned power outage. The same study highlights that the WTP depends users' reliance on electricity, and could therefore be affected by users' adoption of electric heating, cooking, and transportation, as well as off-grid energy storage technologies. This type of study could inform the value chosen for the impact of grid failure for all users. The required size of the financial punishment for rule-breaking could then be calculated.

Using a cost-based reward function may remove the need for a discount function altogether. The horizon for which grid users need to make choices (for example about charging) is short, on the scale of days. The actual administration of the costs takes places on a yearly or at most a monthly basis. Assuming a monthly administration, there should be no perceived difference between the value of an identical cost made today or tomorrow, since the costs accumulate and are only administered by the end of the month. Further research is required to confirm whether users do behave rationally in this sense, and whether a shorter administration period could significantly affect this behaviour.

C. The potential of using game theory to evaluate commons management approaches for STSs

CPR theory originates from analysing SESs, in which natural resources are managed by a community. The step to apply the same theory to STSs is logcial, but not without pitfalls. Existing literature has failed to address the unbuffered nature of resource units in an STS. This thesis suggests that timing of appropriation, rather than the amount appropriated, forms the largest challenge for STSs. This claim is substantiated by highlighting the similarities between electricity infrastructure and road infrastructure, which each suffer from congestion caused by synchronized user behaviour.

Ostrom's design principles for succesfully managed CPRs provide an excellent starting point for the translation of congestion management into game theory. Three key insights allowed for a ruleset in which stability was achieved. The first (1) regards the empowerment of appropriators: appropriators need to have agency, meaning they are in control, but also that they are faced with the consequences of their actions. The second (2) insight regards the functional role of rule-breaking: infractions are required for converging to and maintaining cooperative behaviour. A system designed around a zero-tolerance policy would not be effective, especially if the system relies on voluntary participation. Lastly (3), successful CPR management is only possible if it is facilitated by existing governmental institutions. By assuming the cooperation of the DSO, congestion management approaches can be tested which rely on shared grid information, and a redistribution of existing funds can be suggested in order to create incentives that lead to grid stability.

Similarly, game theoretical simulations provide valuable insights for commons management. The simulation allows for the emergence of various policies, which can be evaluated for desirability. The results showed that some scenarios led to the emergence of a sub-class of players showing aggressive, non-cooperative behaviour. Although the outcome on the operational level was stable, it was also unfair, which can lead to institutional instability on a higher level. The ruleset should be modified to disincentivise undesirable policies.

It is particularly notable that Ostrom's analysis of real-life case studies of succesful CPR management leads to conclusions that are similar to those from Axelrod's game-theoretical analysis of various strategies for cooperation; in both cases the types of behaviour that lead to stable long-term outcomes exhibit high levels of *niceness*, *forgiveness*, and *provocability*. The compatibility of these conclusions from different research methods underscores the potential of using game theory to support the design of institutions for commons management.

Ostrom mentions in her work that the proposed design principles are a work in progress [30]. It is not unthinkable that they have to be adjusted to fit better with STSs, as there are some key differences to SESs. To do so, it is necessary to leave the theoretical world of game theory and to find practical case studies of STSs in which commons management is proven to be successful and stable over time.

D. Policy implications

The literature review at the start of this thesis discusses the existing and proposed measures to address congestion. The analysis of the distribution grid as a CPR and the experimental results from this thesis have highlighted shown the that self-management of grid congestion by users should receive further consideration as a promising solution for congestion management. The policy implications of this are considerable.

A joint publication by net management companies in the Netherlands suggests that an additional 60 - 80.000 kilometers of new electricity cables is necessary by 2050, and requiring up to 4.5 billion euros a year in investments in infrastructure [65]. With regards to residential infrastructure, one in three streets will have to be broken up for expansion works. If true sustainability is the goal, then the environmental impact and opportunity cost of massive infrastructural expansion must be evaluated against alternative solutions. The same money could be used to facilitate and incentivise local congestion management, which would without a doubt be less carbon intensive. The DSOs could offer congested neighborhoods the option between construction work and rising prices or lower prices under self-management.

Ostrom's 7th design principle states that governing bodies must recognize and facilitate self-organisation for it to be effective [30]. Local communities can set up associations to deal with congestion management, which have the potential to be effective in solving a range of energy related issues such as peer-to-peer trading and electricity and heat storage. We have seen that quality grid information is required for self-monitoring and self-management. The DSO could empower local associations by passing through information from smart meters. Hardware could be installed that emulates the grid failure behaviour described in this thesis, where a digital load measurement in the transformer triggers a blackout, and periodically an automatic attempt is made to restore power. Solutions other than a rotation or curtailment system are conceivable, such as a limited number of shared chargers in the neighborhood, for which an appropriation schedule is created.

We have seen that finding a behavioural equilibrium requires a certain amount of training, in which players experience the reward for their actions under various circumstances. To rely on practical experimentation in real life is impractical and would massively disturb the daily life of households. We have also seen that for some rulesets, such as the rotation ruleset, a pre-initialization of beliefs is required for staying in equilbrium. Both for training and pre-initializing beliefs, *serious gaming* could provide an opportunity. Any community that aims to self-manage their local grid could spend some time in a simulated multi-player environment to communally acquire the skills needed to keep the grid stable. Additionally, going through this learning process would allow a community to get a better sense of the the impact that any operational ruleset might have on daily life. Serious gaming is gaining popularity as a tool for co-creating resource management policy [66]. A new role can be created for individuals that specialize in this type of facilitation. This would be a new type of 'green' job, which would likely be easier to fill in the Netherlands than traditional construction jobs, as the Netherlands is experiencing a labour shortage for blue collar workers.

Despite the potential of cooperative approaches, recent EU legislation on unbundling specifies that DSOs should focus on market-based approaches for congestion management [2]. In light of the results of this thesis, this focus can be called short-sighted, as alternative approaches for congestion management have not received adequate attention. EU and national policy will need to allow and encourage DSOs to explore alternatives that are not based on free-market approaches.

E. Future research

This section has already made some suggestions for future research topics that can improve the validity and usefulness of machine learning for incentive evaluation in a game-like scenario. A reward function based on financial cost could expand the parameter space to include external incentives and allow for clearer outcomes in terms of rewards and punishment. A different learning approach that includes an environment model could allow for more complex inputs and behaviours, as well as faster convergence of beliefs and policies.

Additionally, the proposed methodology can be expanded to yield answers about the decentral organisation of more energy-related issues. On the residential level, the inclusion of (electric) heating would increase congestion but also allows for more coordinated behaviour. Domestic PV yield could be included to model a more dynamic electricity cost price, and here too coordination could be beneficial to limit congestion in the opposite direction. Joint use of energy storage technologies could be explored to analyse the outcomes of various proposed rulesets. The scope of the simulation could be expanded from a single feeder to include all feeders that connect to a single transformer.

Lastly, the same methodology could also be applied to a different environment. Grid congestion in the Netherlands is causing major issues for industrial areas, where there is simply not enough capacity to allow new businesses to open. DSOs are pioneering new flexibility trading platforms for large *industrial* users in these areas, but thus far with limited success [21]. These environments are ideal for the application of CPR theory, as industrial users are more actively engaged with their appropriation activities than residential users are, and have to cooperate despite competitive interests.

VI. Conclusion

This thesis aims to add to the existing pool of solutions that exist for addressing congestion on the low-voltage electricity grid. Existing solutions have focused on designing market structures that allow for trading flexibility, or control structures for automated load shifting. However, these solutions face significant barriers to implementation. Therefore, it is necessary to look at the issue from a new perspective.

The **main research question**, which considers the effect of various congestion management rulesets on household behaviour, frames the inquiry into academic literature. The influence of both EU energy policy and infrastructure design on the emergence of congestion is highlighted. Subsequently, theory on common-pool resources (CPR) is introduced as it provides valuable case studies of successful community-based management of shared resources. To support the application of CPR theory to congestion management in a distribution grid, the theoretical differences between socio-ecological systems (SES) and socio-technical system (STS) are addressed. Departing from previous publications, this thesis proposes to define the issue of congestion as time-constrained appropriation of an unbuffered resource.

Subquestion 1 requires an investigation into the potential of game theory to define the issue at hand. An analysis of various game types concludes that community-based congestion management can best be modeled as a stochastic game, in which each household is a player that makes decisions based on grid information and an agreed-upon ruleset. Following from CPR theory, the utility from each player needs to modelled in a way that players are provided with direct feedback on their actions, so that player learning is enabled. **Subquestion 2** is answered by modelling the reward for each player as the amount of energy that is appropriated between decisions. Once game rules are established, it is possible to evaluate what policies players adopt to maximize their reward. In a stochastic game, one way of identifying the equilibrium state in which players have settled on policies that maximize expected reward is by using machine learning. Q-learning and Monte-Carlo simulation are applied to analyse what behaviours emerge in 4 different congestion management scenarios.

In response to **subquestion 3**, the simulation results for the 4 scenarios show what different types of behaviour emerge, and their effect on the functioning of the grid. In the load- and voltage-informed scenarios, players are given grid state information, allowing them to individually adjust their behaviour to avoid grid failure. Here, two types of behaviour emerge: cooperative behaviour of players who decide to wait or stop charging to prevent grid failure, and non-cooperative 'aggressive' behaviour of players who rely on cooperative players to maintain the grid stability. Nonetheless, players are able to significantly reduce congestion, especially when given the higher-quality grid load information.

In the rotation and GridShield scenarios, community cooperation was presumed that enabled the implementation of respectively a charging schedule and a load curtailment infrastructure. From simulations it is apparent that, without incentives for cooperative behaviour, players cannot learn to manage congestion. In the rotation scenario, two parameters are important: the number of groups in the schedule, and the size of the punishment. An optimum balance can be found that maximizes the amount charged while minimizing grid failure. For the GridShield scenario, a significantly smaller incentive can lead to even better outcomes, with almost no grid failure, and very low amount of energy not served at day end.

The resulting punishment values provide an answer to **subquestion 4**, providing specific incentive structures that enable effective congestion management. However, these outcomes are highly tailored to the specific parameters used in the simulation. The parameters, as well as their tested and potential ranges, are displayed in Table VI. The discussion section lists further limitations to the validity of the outcomes, and suggests ways in which more actionable outcomes can be obtained.

Next to the numerical outcomes of the simulations, this research provides more general learnings for designing incentive structures for congestion management in three different dimensions:

- Methodological:
  - A stochastic game can be used to model the interactions of household sharing a low-voltage distribution grid;
  - Q-learning can be used as a machine learning technique to evaluate the incentive structures in an STS such a distribution grid, but has significant limitations.
- Empirical:
  - Self-interested players sharing a low-voltage distribution grid can learn to adequately manage congestion, if provided with high quality information about the state of the grid;
  - Congestion is reduced significantly if players organise using a rotating charging schedule or implement automatic load curtailment (GridShield), provided that appropriately sized sanctions for infractions are implemented;
  - The incentive evaluation demonstrates that, in the setup tested, automatic load curtailment (GridShield) requires 1/5th of the punishment size of the rotation schedule to address congestion.
- Socio-technical:
  - STS differ from SES in that they generally provide an unbuffered, time-constrained resource, and consequently they are threatened by excess demand at a single point in time rather than excess demand over a period of time;
  - Approaching the grid congestion issue from a commons perspective provides valuable design guides for voluntary congestion management, namely that:
    - ∗ The incentive structure must be setup in such a way that grid users are faced with the consequences of their actions so that learning is enabled;
    - ∗ Rule-breaking must be integrated as an essential behaviour for converging to and maintaining cooperative behaviour;
    - ∗ Overarching institutions such as DSOs must facilitate congestion management by recognizing local associations, providing grid data, and redistributing funds to create incentives.
  - Because the behaviour of players in a stochastic game is dynamic, the initialization of beliefs matters. The incentive to adhering to a ruleset only exists if other players are already adhering to that ruleset. If the beliefs are not initialized at the desired equilibrium, then that equilibrium may never be reached.

This work suggests the development of further research for each of these dimensions. The methodology could be improved by developing a utility function based on financial cost. A model-based learning approach could be implemented to overcome the issues of Q-learning. In terms of empirical conclusions, the same methodology could be applied to a large variety of congestion scenarios, such as various residential and industrial energy grids. The methodology could conceivable also be applied to other types of congestion, such as traffic congestion. Lastly, further analysis can be done on a socio-technical level to support the practical implementation of commons-based congestion management, and to compare this solution to other (top-down) approaches.

In this work, the nature of congestion has been discussed and analysed. As a result of the energy transition, congestion is a rapidly growing problem that is likely here to stay. In light of billion-dollar investments and sweeping legislative changes, it is paramount that solutions from every conceivable angle are considered.

I would like to use these final sentences to urge a warning to analysts who believe that with enough investments into infrastructure and technology congestion can be resolved. A brief look into the adjacent field of traffic engineering shows that this is a Sisyphean task. If the aim of the energy transition is to achieve true sustainability, then policy makers must consider approaches that limit the need for acquiring and constructing additional power and/or IT infrastructure. A congestion management approach based on voluntary cooperation can provide the foundation for a societal shift towards a more conscious use of energy.

Name	Symbol	Tested range	Possible range	Effect on convergence	Effect on outcomes
Decision chance	p_c	$1/(2N)$	$0-1$	Higher decision chance increases convergence speed	More simultaneous decisions reduces grid stability
Discount rate	γ	0.95	$0-1$	Lower discount rate reduces clarity of rewards	Lower discount rate reduces grid stability
Learning rate	α	Exponential decay Fixed at 0.1	$0-1$	Affects convergence speed and accuracy	-
Exploration rate	ε	Exponential decay Fixed at 0.1	$0-1$	Determines stability of behaviour and consequently the clarity of of rewards	High exploration rate reduces grid stability
Provided state information	S	Voltage Load Rotation Grid failure		Provided information determines what players can learn	Quality information allows players to maintain grid stability
Action set	A	Charge Wait Charge (GridShield)		Larger action set reduces clarity of rewards and convergence speed	Larger action set allows for more tailored behaviour
Number of players	N	26	$0-\infty$	A more crowded grid requires more behavioural adjustments, reducing convergence speed	A more crowded grid will reduce grid stability and increase unserved demand
Grid limit	R_{max}	100	$0-\infty$		
EV penetration		100%	$0-100\%$		
Negative reward for grid failure		0	$0-\infty$	-	Higher impact of failure increases grid stability
Simulation range in minutes		$1000-1439$	$0-1439$	Broader range reduces clarity of rewards	-
			Rotation		
Punishment		$50-300$	$0-\infty$	Punishment of 190 or higher required for convergence	Trade-off between unserved demand and grid stability
Number of groups		$2-8$	$1-N$	-	4 or more groups required for grid stability. Higher number of groups increases unserved EV demand.
			GridShield		
Punishment		$0-300$	$0-\infty$	Punishment of 35 or higher required for convergence	Punishment over 35 allows for perfect outcomes

Table VI: Summary of parameters

References

[1] European Commission. Directorate General for Communication., *REPowerEU with clean energy.* eng. LU: Publications Office, 2022.

[2] M. Buchmann, "How decentralization drives a change of the institutional framework on the distribution grid level in the electricity sector – The case of local congestion markets", en, *Energy Policy*, vol. 145, p. 111 725, Oct. 2020, ISSN: 0301-4215. DOI: 10.1016/j.enpol.2020.111725.

[3] *Hoeveel stroom gebruikt een volledig elektrisch huishouden?*, en-US, Feb. 2017.

[4] A. Kättlitz, P. Dante, G. Saunders, *et al.*, "TYNDP 2022 Scenario Report — Version. April 2022", en, ENTSOG; ENTSO-E, Tech. Rep., Apr. 2022, p. 78.

[5] *EU deal to end sale of new CO2 emitting cars by 2035*, en, Text, Oct. 2022.

[6] G. Hoogsteen, "A Cyber-Physical Systems Perspective on Decentralized Energy Management", English, Dec. 2017. DOI: 10.3990/1.9789036544320.

[7] *StatLine - Zonnestroom; vermogen en vermogensklasse, bedrijven en woningen, regio*, nl, Dec. 2022.

[8] *Capaciteitskaart invoeding elektriciteitsnet*, nl.

[9] *Sector PMR Indicators - OECD Indicators in Excel*, Sep. 2022.

[10] M. G. Pollitt, "The European Single Market in Electricity: An Economic Assessment", en, *Review of Industrial Organization*, vol. 55, no. 1, pp. 63–87, Aug. 2019, ISSN: 1573-7160. DOI: 10.1007/s11151-019-09682-w.

[11] O. Celasun, A. Mineshima, N. Arregui, *et al.*, "Surging Energy Prices in Europe in the Aftermath of the War: How to Support the Vulnerable and Speed up the Transition Away from Fossil Fuels", en, *IMF Working Papers*, vol. 2022, no. 152, p. 1, Jul. 2022, ISSN: 1018-5941. DOI: 10.5089/9798400214592.001.

[12] *The Harmonised Electricity Role Model*, Mar. 2022.

[13] L. Hirth and S. Glismann, "Congestion Management: From Physics to Regulatory Instruments", eng, Kiel, Hamburg: ZBW – Leibniz Information Centre for Economics, Working Paper, 2018.

[14] S. Rusck, "The simultaneous demand in distribution network supplying domestic consumers", *ASEA journal*, vol. 10, no. 11, pp. 59–61, 1956.

[15] P. to Phase, *Netten voor distributie van elektriciteit, Phase to Phase*, nl-NL, 3rd ed. 2020, ISBN: 978-90-817983-1-0.

[16] N. Shabbir, L. Kutt, V. Astapov, *et al.*, "Congestion control strategies for increased renewable penetration of photovoltaic in LV distribution networks", en, *Energy Reports*, vol. 8, pp. 217–223, Dec. 2022, ISSN: 2352-4847. DOI: 10.1016/j.egyr.2022.10.184.

[17] E. Mulenga, M. H. J. Bollen, and N. Etherden, "A review of hosting capacity quantification methods for photovoltaics in low-voltage distribution grids", en, *International Journal of Electrical Power & Energy Systems*, vol. 115, p. 105 445, Feb. 2020, ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2019.105445.

[18] Y. Yu, D. Reihs, S. Wagh, *et al.*, "Data-Driven Study of Low Voltage Distribution Grid Behaviour With Increasing Electric Vehicle Penetration", *IEEE Access*, vol. 10, pp. 6053–6070, 2022, Conference Name: IEEE Access, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3140162.

[19] P. Bach Andersen, J. Hu, and K. Heussen, "Coordination strategies for distribution grid congestion management in a multi-actor, multi-objective setting", in *2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, ISSN: 2165-4824, Oct. 2012, pp. 1–8. DOI: 10.1109/ISGTEurope.2012.6465853.

[20] F. Tangerding, I. A. M. Varenhorst, G. Hoogsteen, M. E. T. Gerards, and J. L. Hurink, "GridShield: A Robust Fall-Back Control Mechanism for Congestion Management in Distribution Grids", in *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Oct. 2022, pp. 1–5. DOI: 10.1109/ISGT-Europe54678.2022.9960301.

[21] T. Schittekatte and L. Meeus, "Flexibility markets: Q&A with project pioneers", en, *Utilities Policy*, vol. 63, p. 101 017, Apr. 2020, ISSN: 0957-1787. DOI: 10.1016/j.jup.2020.101017.

[22] European Parliament and Council of the European Union, *Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU*, Jun. 2019.

[23] R. A. Verzijlbergh, L. J. De Vries, G. P. J. Dijkema, and P. M. Herder, "Institutional challenges caused by the integration of renewable energy sources in the European electricity sector", en, *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 660–667, Aug. 2017, ISSN: 1364-0321. DOI: 10.1016/j.rser.2016.11.039.

[24] J. Katz, "Linking meters and markets: Roles and incentives to support a flexible demand side", en, *Utilities Policy*, vol. 31, pp. 74–84, Dec. 2014, ISSN: 0957-1787. DOI: 10.1016/j.jup.2014.08.003.

[25] *Innovation Outlook: Smart charging for electric vehicles*, en. Abu Dhabi: International Renewable Energy Agency, May 2019, ISBN: 978-92-9260-124-9.

[26] R. A. Verzijlbergh, L. J. De Vries, and Z. Lukszo, "Renewable Energy Sources and Responsive Demand. Do We Need Congestion Management in the Distribution Grid?", *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2119–2128, Sep. 2014, Conference Name: IEEE Transactions on Power Systems, ISSN: 1558-0679. DOI: 10.1109/TPWRS.2014.2300941.

[27] X. Li and A. Jenn, "Energy, Emissions, and Cost Impacts of Charging Price Strategies for Electric Vehicles", *Environmental Science & Technology*, vol. 56, no. 9, pp. 5724–5733, May 2022, Publisher: American

Chemical Society, ISSN: 0013-936X. DOI: 10.1021/acs.est.1c06231.

[28] R. Hennig, M. Jonker, S. Tindemans, and L. de Vries, “Capacity Subscription Tariffs for Electricity Distribution Networks: Design Choices and Congestion Management”, in *2020 17th International Conference on the European Energy Market (EEM)*, ISSN: 2165-4093, Sep. 2020, pp. 1–6. DOI: 10.1109/EEM49802.2020.9221994.

[29] K. Zhou and L. Cai, “Randomized PHEV Charging Under Distribution Grid Constraints”, *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 879–887, Mar. 2014, Conference Name: IEEE Transactions on Smart Grid, ISSN: 1949-3061. DOI: 10.1109/TSG.2013.2293733.

[30] E. Ostrom, *Governing the commons: the evolution of institutions for collective action* (Canto classics). Cambridge, United Kingdom: Cambridge University Press, 2015, OCLC: ocn911061917, ISBN: 978-1-107-56978-2 978-1-316-42393-6.

[31] G. Hardin, “The Tragedy of the Commons”, *Science*, vol. 162, no. 3859, pp. 1243–1248, Dec. 1968, Publisher: American Association for the Advancement of Science. DOI: 10.1126/science.162.3859.1243.

[32] E. Ostrom, “Tragedy of the Commons”, English, in *The New Palgrave Dictionary of Economics*, Second, New York: Palgrave Macmillan, 2008, ISBN: 978-0-333-78676-5.

[33] C. Acosta, M. Ortega, T. Bunsen, B. P. Koirala, and A. Ghorbani, “Facilitating Energy Transition through Energy Commons: An Application of Socio-Ecological Systems Framework for Integrated Community Energy Systems”, en, *Sustainability*, vol. 10, no. 2, p. 366, Feb. 2018, Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2071-1050. DOI: 10.3390/su10020366.

[34] S. Eslamizadeh, A. Ghorbani, and M. Weijnen, “Establishing industrial community energy systems: Simulating the role of institutional designs and societal attributes”, en, *Journal of Cleaner Production*, vol. 419, p. 138 009, Sep. 2023, ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2023.138009.

[35] N. Y. Soltani, S.-J. Kim, and G. B. Giannakis, “Real-Time Load Elasticity Tracking and Pricing for Electric Vehicle Charging”, *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1303–1313, May 2015, Conference Name: IEEE Transactions on Smart Grid, ISSN: 1949-3061. DOI: 10.1109/TSG.2014.2363837.

[36] *Profielen electriciteit 2023 (nieuwe stijl) v1.00*, nl-NL, May 2022.

[37] *NDW Dexter*.

[38] D. Braess, A. Nagurney, and T. Wakolbinger, “On a Paradox of Traffic Planning”, en, *Transportation Science*, vol. 39, no. 4, pp. 446–450, Nov. 2005, ISSN: 0041-1655, 1526-5447. DOI: 10.1287/trsc.1050.0127.

[39] D. Witthaut and M. Timme, “Braess’s paradox in oscillator networks, desynchronization and power outage”, en, *New Journal of Physics*, vol. 14, no. 8, p. 083 036, Aug. 2012, Publisher: IOP Publishing, ISSN: 1367-2630. DOI: 10.1088/1367-2630/14/8/083036.

[40] B. Wolterman, T. Duurkoop, M. Zweistra, E. Hiep, and M. van Biezen, “Nationaal Laadonderzoek 2022”, nl, Rijksdienst voor Ondernemend Nederland, Tech. Rep., Sep. 2022, p. 45.

[41] M. Mogridge, D. Holden, J. Bird, and G. Terzis, “The Downs/Thomson Paradox and the Transportation Planning Process”, *International Journal of Transport Economics / Rivista internazionale di economia dei trasporti*, vol. 14, no. 3, pp. 283–311, 1987, Publisher: Accademia Editoriale, ISSN: 0303-5247.

[42] P. Kuss and K. A. Nicholas, “A dozen effective interventions to reduce car use in European cities: Lessons learned from a meta-analysis and transition management”, en, *Case Studies on Transport Policy*, vol. 10, no. 3, pp. 1494–1513, Sep. 2022, ISSN: 2213-624X. DOI: 10.1016/j.cstp.2022.02.001.

[43] S. Ribeiro, M. Figueroa, F. Creutzig, C. Dubeux, J. Hupe, and S. Kobayashi, “Global Energy Assessment - Toward a Sustainable Future. Chapter 9 - Energy End-Use: Transport”, in Jun. 2012, ISBN: 978-1-107-00519-8.

[44] *Stroomaansluiting & gasaansluiting in de meterkast - Gaslicht.com*, nl.

[45] A. Nandram, *Maandag en vrijdag thuis, dinsdag en donderdag op kantoor: Thuiswerk ontregelt NS*, nl-NL, Section: Topverhalen vandaag, Feb. 2023.

[46] D. Fudenberg and J. Tirole, *Game theory*. Cambridge, Mass: MIT Press, 1991, ISBN: 978-0-262-06141-4.

[47] J. Von Neumann and O. Morgenstern, *Theory of games and economic behavior* (Princeton classic editions), 60th anniversary ed. Princeton, N.J. ; Woodstock: Princeton University Press, 2007, OCLC: ocm78989126, ISBN: 978-0-691-13061-3.

[48] J. Nash, “Non-Cooperative Games”, *Annals of Mathematics*, vol. 54, no. 2, pp. 286–295, 1951, Publisher: Annals of Mathematics, ISSN: 0003-486X. DOI: 10.2307/1969529.

[49] R. Axelrod, “Effective choice in the prisoner’s dilemma”, *Journal of conflict resolution*, vol. 24, no. 1, pp. 3–25, 1980, Publisher: Sage Publications Sage CA: Los Angeles, CA.

[50] R. Axelrod, “More Effective Choice in the Prisoner’s Dilemma”, *The Journal of Conflict Resolution*, vol. 24, no. 3, pp. 379–403, 1980, Publisher: Sage Publications, Inc., ISSN: 0022-0027.

[51] R. Axelrod and R. O. Keohane, “Achieving Cooperation under Anarchy: Strategies and Institutions”, *World Politics*, vol. 38, no. 1, pp. 226–254, 1985, Publisher: Cambridge University Press, ISSN: 0043-8871. DOI: 10.2307/2010357.

[52] S. Abapour, M. Nazari-Heris, B. Mohammadi-Ivatloo, and M. Tarafdar Hagh, "Game Theory Approaches for the Solution of Power System Problems: A Comprehensive Review", en, *Archives of Computational Methods in Engineering*, vol. 27, no. 1, pp. 81–103, Jan. 2020, ISSN: 1886-1784. DOI: 10.1007/s11831-018-9299-7.

[53] L. S. Shapley, "Stochastic Games", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 39, no. 10, pp. 1095–1100, Oct. 1953, ISSN: 0027-8424.

[54] M. Bowling and M. Veloso, "An analysis of stochastic game theory for multiagent reinforcement learning", Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science, Tech. Rep., 2000.

[55] C. B. Kobus, R. Mugge, and J. P. Schoormans, "Washing when the sun is shining! How users interact with a household energy management system", *Ergonomics*, vol. 56, no. 3, pp. 451–462, Mar. 2013, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00140139.2012.721522, ISSN: 0014-0139. DOI: 10.1080/00140139.2012.721522.

[56] E. Dogan, J. Bolderdijk, and L. Steg, "Making Small Numbers Count: Environmental and Financial Feedback in Promoting Eco-driving Behaviours", *Journal of Consumer Policy*, vol. 37, no. 3, pp. 413–422, Sep. 2014, Publisher: Springer Nature, ISSN: 01687034. DOI: 10.1007/s10603-014-9259-z.

[57] B. Nijenhuis, *Electric vehicle charging in the Dutch low-voltage grid*, Aug. 2020.

[58] L. Thurner, A. Scheidler, F. Schäfer, *et al.*, "Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems", *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–6521, Nov. 2018, Conference Name: IEEE Transactions on Power Systems, ISSN: 1558-0679. DOI: 10.1109/TPWRS.2018.2829021.

[59] *Huishoudens; samenstelling, grootte, regio, 1 januari*, nl-NL, webpagina, Last Modified: 2022-08-24T02:00:00+02:00, Aug. 2022.

[60] C. J. C. H. Watkins and P. Dayan, "Q-learning", en, *Machine Learning*, vol. 8, no. 3, pp. 279–292, May 1992, ISSN: 1573-0565. DOI: 10.1007/BF00992698.

[61] F. Charbonnier, T. Morstyn, and M. D. McCulloch, "Scalable multi-agent reinforcement learning for distributed control of residential energy flexibility", en, *Applied Energy*, vol. 314, p. 118 825, May 2022, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2022.118825.

[62] V. Kreinovich and C. Stylios, "Why Fuzzy Cognitive Maps Are Efficient", *International Journal of Computers Communications & Control*, vol. 10, p. 65, Oct. 2015. DOI: 10.15837/ijccc.2015.6.2073.

[63] O. T. Eldakar, "Hawk-Dove Model, The", en, in *Encyclopedia of Personality and Individual Differences*, V. Zeigler-Hill and T. K. Shackelford, Eds., Cham: Springer International Publishing, 2020, pp. 1905–1910, ISBN: 978-3-319-24612-3. DOI: 10.1007/978-3-319-24612-3_1645.

[64] F. Carlsson, M. Kataria, E. Lampi, and P. Martinsson, "Past and present outage costs – A follow-up study of households' willingness to pay to avoid power outages", en, *Resource and Energy Economics*, vol. 64, p. 101 216, May 2021, ISSN: 0928-7655. DOI: 10.1016/j.reseneeco.2021.101216.

[65] "Samenvatting Het Energiesysteem van de toekomst.pdf", Netbeheer Nederland, Tech. Rep., Jun. 2023.

[66] P. Edwards, L. Sharma-Wallace, A. Wreford, *et al.*, "Tools for adaptive governance for complex social-ecological systems: A review of role-playing-games as serious games at the community-policy interface", en, *Environmental Research Letters*, vol. 14, no. 11, p. 113 002, Oct. 2019, Publisher: IOP Publishing, ISSN: 1748-9326. DOI: 10.1088/1748-9326/ab4036.