Cyber-Physical Systems Design:
Electricity Markets and Network Security

by

Mohammad Rasouli

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in the University of Michigan
2018

Doctoral Committee:

Professor Demosthenis Teneketzis, Chair
Assistant Professor Mohammad Akbarpour, Stanford University
Professor Ian Hiskens
Professor William Hogan, Harvard University
Assistant Professor Shaowei Ke
Professor Paul Milgrom, Stanford University
Professor Asuman Ozdaglar, Massachusetts Institute of Technology

Mohammad Rasouli
rasouli@umich.edu
ORCID iD: 0000-0002-4038-1456

To those who have faith in the possibility of a world where every human has the opportunity to follow her/his interest and flourish her/his talent, a world where no individual is deprived from the basics of a descent life, and a world where people are in cooperation to grow not at war to destroy. Such world will come true only by the efforts and contributions of those who believe in it.

ACKNOWLEDGMENTS

I want to thank Demos for being a great adviser. In our phone conversation before joining the PhD program, he said he will support me to follow my research interests in the intersection of engineering and economics, and he did so. Demos was patient with my education, gave me freedom to work on topics I want and how I want, and continuously encouraged me through this research. He taught us to follow researchers who conduct fundamental and valuable research even if they are John Nash! Demos and Barbara were great friends. We had long conversations on wide set of topics from politics and sociology to art and ethics. They genuinely shared their experiences with us. Thanks Demos and Barbara.

I am fortunate that I had my research role models in my thesis committee. Thanks to Paul Milgrom for accepting me in his office and at his house and listening carefully to my ideas. His writings shaped my view to these problems and his encouragement to follow this line of research were very inspiring to me. Thanks to William Hogan who has indirectly advised me through his papers and has shown me how a theory research can address major real-world policy challenges. He read my work carefully, accepted me at his house kindly and shared his experiences with me generously. Thanks to Asuman Ozdaglar who was with this research during our joint grant and never stopped sharing her advice with us whenever possible. Thanks to Ian Hiskens whose course on electricity markets in my first semester shaped the origin of this research. His door was always open to me for questions. Thanks to Mohammad Akbarpour for our 12+ years of friendship back to high school and college years. We went to travels and participated in students activities together. I always enjoy discussing with him almost on all topics and admire his out of the box views. And thanks to Shaowei Ke for accepting me at the Econ Theory Lunch seminars. I enjoyed his deep view of the problems and his accurate comments.

I am also fortunate that my family and part of my extended family were with me during my PhD. I would like to thank my parents, also my life-long teachers, Hojjat and Mahasti. They made the best year of my life by coming to Ann Arbor as part of my fathers sabbatical at the University of Michigan Near Eastern Studies Department. We went to road trips together in the East Coast and West Coasts, Grand Canyon and Zion Mountain, Yellowstone and North Dakota. My mother taught my friends and me how to think about complicated social problems in her sociology class, and my father introduced us the beauty of Persian and Arabic literature in our weekly poetry gathering. In all my life, my parents where available to talk to me on all topics and these discussions formed my view to the life. Thank you and I love you. Also thanks to Rezvan, a great sister and a great Chicago artist, who showed me there exists a different world in parallel to our engineering world in which people think about different topics and in a different way. Rezvan has always been a great support for me, and stood beside me in the hardest times. Thank you Rezvan. Thanks to my grandfather, Abbas, who joined us in our long trip to East Coast and in our trip to Michigan Upper Peninsula. Also to my great uncles Vamegh and Mohammad who came to visit me in Ann Arbor. They all truly want my success and happiness and are always available for support.

I have had the opportunity to learn from great teachers in my life. I owe to Jamal Golestani who changed me from a startup engineer to a theory research PhD student by showing me the beauty of the mathematical models. Ever since he has been a true adviser for me. I also owe to my teachers in Allame Helli high school without whose efforts I could not reach this point. Grand Master Chong, Master Busch and Master Sahba have been great teachers for me at the University of Michigan Taekwondo Club. They helped me advance in the art of Taekwondo. The club has always been inclusive and diverse. I formed great friendships in the club, we practiced together, traveled together, and won the ECTC Tournament Second Division together. Go Blue UMTKD!

I have been fortunate to have the company of great friends in my life. Erik, Farzane, Hamid and Ouyang were friends who were also my groupmates and coauthors. Other (random) names out of many, Mohammad, Navid, Javad, Ali, Hossein, Amir, Ahmad, Mina, Hadieh, Niusha, Haidar, Mehrzad, I am grateful for the time we had together. We read poems of Rumi and Hafez and studied philosophy, we played soccer and traveled, we watched movies and had gatherings. Thank you guys and may our friendships last forever.

TABLE OF CONTENTS

Dedication . . . . **ii**

Acknowledgments . . . . **iii**

List of Figures . . . . **vii**

List of Appendices . . . . **viii**

Abstract . . . . **ix**

Chapter

1 Introduction **1**

1.1 Motivation 1
1.2 Contribution of the Thesis 5

2 Markets with Complex Policy Constraints: Sustainable, Reliable and Price Efficient Electricity **8**

2.1 Introduction 8
2.1.1 Literature Review 14
2.2 Model 18
2.3 Efficient Auction of a Divisible Good 24
2.3.1 Model: Electricity Spot Markets 25
2.3.2 Design/Mechanism 25
2.3.3 Interpretation of the Design 27
2.3.4 Comparison of the Elastic and Inelastic Static Auctions 28
2.3.5 Properties 28
2.4 Efficient Auction with Individual Constraints 31
2.4.1 Model: Investment without Reliability Constraint 31
2.4.2 Design/Mechanism 32
2.4.3 Properties 32
2.4.4 Interpretation 33
2.5 Efficient Auction with Homogeneous (Additive) Joint Constraints 33
2.5.1 Model: Electricity Investment with Reliability Constraint 34
2.5.2 Design/Mechanism 34
2.5.3 Properties 35
2.6 Long-Term Interaction with Common Uncertainty 37
2.6.1 Model: Long Term Investment with Reliability Constraints 37

2.6.2 Design/Mechanism . . . 38
2.6.3 Properties . . . 38
2.7 Implementing Electricity Policy Targets . . . 41
2.7.1 Design/Mechanism . . . 41
2.7.2 Contribution to the Electricity Policy Debates . . . 45
2.8 Conclusion and Future Directions . . . 46

3 Markets with Complex Physical Constraints: Networked Electricity Spot Market . . . 49

3.1 Introduction . . . 49
3.1.1 Existing Approaches for Electricity Markets with Networks . . . 50
3.1.2 Contribution . . . 51
3.2 Model . . . 51
3.3 Elements of Networked Markets . . . 52
3.4 Market Design for Electricity Trade over Network . . . 54
3.4.1 The Centralized Problem and The Associated KKT Conditions . . . 55
3.4.2 The Local Public Good Mechanism . . . 56
3.4.3 Properties of the Mechanism . . . 57
3.5 Future Directions . . . 63

4 Large-Scale CPS: Network Security with Non-Bayesian Uncertainty . . . 64

4.1 Introduction . . . 64
4.1.1 Organization of the Chapter . . . 65
4.1.2 Notation . . . 66
4.2 The Security Model . . . 67
4.3 The Defense Problem . . . 69
4.3.1 Information State . . . 70
4.3.2 Sequential Decomposition and Dynamic Programming . . . 72
4.4 Approximation to the Defense Problem . . . 73
4.4.1 Local Defense Problems . . . 73
4.4.2 Approximating the Local Defense Problems . . . 78
4.4.3 Scalability . . . 80
4.5 Example . . . 81
4.6 Conclusion and Future Directions . . . 86

5 Conclusion and Future Directions . . . 88

Bibliography . . . 90

Appendices . . . 98

LIST OF FIGURES

2.1 Scarcity price: When inelastic demand exceeds capacity, market fails to determine scarcity price. 10
2.2 Scarcity price: With elastic demand, market determines scarcity price at peak load. . . 10
2.3 Timeline of events in the electricity industry during period t to $t+1$. 20
2.4 Timeline of events in year t for the proposed markets of Section (2.6.2). 39
2.5 Timeline for the proposed capacity-and-energy mechanism 43

4.1 Timeline of events for a given time-step. 68
4.2 Information History of the Defender 70
4.3 Timeline of events for the local defense problem (P_k) for a given time-step. 74
4.4 Repeated application of the mincut algorithm to obtain three clusters, $\mathscr{N}_1 = \{3,4\}$, $\mathscr{N}_2 = \{2\}$, and $\mathscr{N}_3 = \{1,5\}$. 84

LIST OF APPENDICES

A Proof of Properties in Section (2.3) 98

B Proof of Properties in Section (2.4) 105

C Proof of Properties in Section (2.5) 111

D Extension of the Auctions with Individual Constraints in Section (2.4) to Long-term Interaction of GenCos with Symmetric Uncertainty 121

ABSTRACT

This thesis presents *Cyber-Physical Systems Design (CPS Design)*. Design of CPS is challenging and requires interdisciplinary studies of engineering and economics because of the distinguishing features of CPS: strategic (self profit-maximizing) decision makers, complex physical constraints, and large-scale networked systems. We study these features by focusing on *designing markets with complex constraints* including both policy and physical constraints, and *decomposing large-scale CPS* within the context of electricity markets and network security.

We first study market design for implementation of complex electricity policy targets, i.e. *sustainability*, *reliability*, and *price efficiency*, by efficient design of spot, carbon, and capacity markets that correct the deficiencies of the current electricity markets; this design does not take into account the network constraints due to the Kirchhoff's laws. To address this problem, we develop a framework based on the design of *efficient auctions with constraints*. Our market design sheds light on major debates in electricity policy including capacity-and-energy vs energy-only markets, carbon market vs carbon tax, and use of price or offer caps.

Second, we add network constraints due to Kirchhoff's laws of current and voltage, which are unique to electricity networks, to the design of electricity spot markets with complex physical constraints. To address this problem, we develop a framework for the design of *networked markets* based on the ideas from local public goods.

Finally, we study the design of defense policies for large-scale network security. Our approach is to design *approximately optimal* defense policies that are computable. We develop a framework based on the notion of *influence graph*, which captures the connectivity of the security states of the system elements, to decompose the system into subsystems. We then design approximately

optimal defense policies for each sub-system. We consider non-Bayesian uncertainty and even though we do not model the attacker as a strategic decision maker, we compensate (in part) for the lack of this feature by adopting a minmax performance criterion.

CHAPTER 1

Introduction

1.1 Motivation

This thesis presents *Cybe-Physical Systems Design (CPS Design)*. CPS are systems that consist of a physical layer (such as lines/routes, sensors, actuators) and a cyber layer (such as communication protocols, information system, and algorithms). Examples of these systems are the electricity grid, security systems, transportation networks, water network and smart devices (Internet of Things). CPS play a central and ever increasing role in our modern life. For these systems, the design of the cyber and physical layer must be done in a coordinated manner.

Designing cyber-physical systems is challenging and requires interdisciplinary studies because of the following features: the presence of strategic decision makers, the existence of complex physical constraints, and the large-scale nature of CPS (Milgrom, 2017). We describe these features below.

A decision maker is strategic if he/she considers how his/her actions will affect the others' actions and the environment. The system operator cannot prescribe the actions to strategic decision makers directly. Instead, he should *design* a mechanism to provide correct incentives for them so that they take the actions desired by him. For example, in the restructured electricity industry, strategic generation companies (GenCos) want to increase their profit by under-producing electricity and selling it to the demand at high prices. The system operator in this case should design appropriate electricity markets that provide correct incentives/punishments for GenCos in order to sell electricity at the socially optimal amount and at the efficient price. Strategic decision makers and mechanism design are studied primarily in economics (Börgers, 2015; Vohra, 2011).

Physics of CPS impose complex constraints in their design. For example, the flow of power in electricity networks is governed by Kirchhoff's laws; the spread of an attack in a communication network is determined by the underlying connectivity of the nodes; the reliability constraints for electricity imposes a very high bar on the minimum amount of investment in electricity capacity. Complex physical constraints make CPS different from existing models in engineering or economics and require new design methods. For example, physical constraints can prevent a perfectly competitive market from occurring and necessitate *market design* (Milgrom, 2017). To see this, consider congested transmission lines in electricity networks. Their congestion can result in local markets with local market power, and the hard constraint on maximum carbon emission in producing electricity can result in a public good problem. As a result, electricity markets are not perfectly competitive; the system operator should design markets by incorporating prices in an effective way while still maintaining enough direct control to ensure that complex constraints are satisfied.

Finally, CPS are large-scale in the sense that the amount of computation or communication required for optimal operation of these systems is beyond the computational power of existing computers or the communication capacity of existing links. For example consider security of the Internet. Because of the limited capacity of communication links, it is not possible for any single node in this network to know the security state of all the other nodes, and even if it does, using this information for calculating an optimal defense policy is computationally infeasible. The computation and communication limits enforce a paradigm shift in designing CPS; rather than optimal performance, the system operator should look for *good enough* performance (Akbarpour and Nikzad, 2017; Kleinberg and Tardos, 2006; Milgrom and Segal, 2017). The resulting design should allow for simple and fast enough decision making. Good enough performance of large-scale CPS can be achieved by using heuristics and decomposing the system into small sub-systems the size of which is compatible with existing computation and communication capabilities.

Networks play a central role in the design of CPS with the above features, because they can be used both to model complex physical constraints and to capture the connectivity of the elements in a large-scale CPS (Akbarpour et al., 2017; Easley and Kleinberg, 2010).

Other than the above distinguishing features, CPS have two additional major features that are in common with the earlier man-made systems: dynamic decision making, and uncertainty (Kumar and Varaiya, 1986). Examples of dynamic CPS are long-term interactions among GenCos in electricity markets, or dynamic spread of attacks in a network security. Examples of uncertainty in electricity industry are the uncertain oil price and sudden breakthroughs in renewable technologies. In security systems uncertainties are due to imperfect observation of the security state or uncertainty about attacker's information and strategy.

Design of cyber-physical systems with the above features is the main focus/topic of this thesis. We address the *design of markets with complex constraints* and *decomposition of large-scale CPS* by concentrating on two applications: electricity markets and network security. First, we study designing markets with complex policy constraints by focusing on designing electricity markets that implement electricity policy targets i.e. *sustainability*, *reliability* and *price efficiency* (Chapter 2). These policy targets and the physical constraints in electricity industry are in the form of either individual or homogeneous (additive) joint constraints. We develop a framework of *efficient auctions* *with constraints* to achieve our market design goal. Next, we study designing markets with complex physical constraints by adding Kirchhoff's laws of current and voltage (network constraints) to the already existing constraints and design electricity spot markets (Chapter 3). We develop a framework of *efficient networked markets* based on ideas from local public good markets. Our market design allows for efficient (and scalable) dispatch of electricity. Finally, we study large-scale CPS by focusing on the defense of large-scale network security (Chapter 4). We use the notion of *influence graph* that captures the connectivity of the nodes, to form sub-systems by clustering the graph, and to determine a defense policy for each sub-system by appropriate mapping of the information space.

Electricity policies in most countries target providing reliable and sustainable electricity with efficient price to customers (IEA, 2016; Ritzenhofen et al., 2014; UNDP, 2015). Reliability refers to meeting the uncertain real-time demand almost at all times (one blackout every ten years), sustainability refers to overcoming the global warming, and efficient price[1] (affordablity) refers to prices close to marginal cost of production. System operators translate reliability target into *planning reserve requirement* that is the total invested capacity should exceed the maximum possible demand by a reserve margin (%15 to %20). Sustainability target is also translated into *carbon cap* which is a cap for total carbon emission. Addressing the policy targets for the electricity industry should consider both the incentive constraints of the strategic GenCos and the physical constraints of the electricity grid. Example of the physical constraints are ramping constraints, lines thermal capacity constraints, non-convexity of the production cost function, Kirchhoff's laws of current and voltage, etc. Physical constraints make electricity market far from perfect competition and necessitate the design of special markets for electricity. Among the physical constraints, the most challenging are Kirchhoff's laws (loop flow effect) which make electricity network different from any other CPS networks; loop flow effect connects the production of all nodes to each other and introduces externalities in a unique form.

Current electricity markets suffer from underinvestment, inefficient mechanisms for limiting carbon emission and high prices. There are major debates on the correct solutions for these problems: The first is centered around the need for non-market mechanisms like price-cap, offer-cap and market-monitoring so as to achieve efficient prices. The second is about the need for mechanisms with direct capacity incentives, like capacity subsidy or capacity market, so as to achieve the reliability target; this approach is referred to as *energy-and-capacity* solution in contrast to *energy-only* solution. Even without planning reserve requirements, some references argue that capacity mechanisms are required because GenCos' investment will not be socially optimal due to their market power (Garcia and Shen, 2010; Murphy and Smeers, 2005; Novy-Marx, 2007). The third

[1]By prices we refer to the prices in the wholesale market, not the regulated price in the end-consumers monthly bills.

is with regard to carbon tax that is introduced so as to achieve sustainability target and refers to the need for the system operator to set the carbon tax (price policy) as opposed to the possibility of finding the efficient carbon tax by carbon markets (quantity policy). On the sustainability topic, there is also an ongoing debate tied to political issues on charging carbon vs. supporting renewables for reducing carbon emission. While the above debates are still ongoing, the electricity industry is under major technological changes, with the most prominent ones being smart grids that introduce demand response and elastic demand, and renewable technologies that introduce more uncertainty and lower prices in the electricity spot market.

The existing studies of electricity policies are limited to addressing reliability, sustainability and price efficiency separately and in isolation; however these issues are in fact intertwined (Kök et al., 2016). For example, any changes in the spot market for affordable prices, like price caps, can reduce reliability by reducing investment incentives (Roques and Savva, 2009); reliability subsidies, like capacity investment subsidy, work against carbon taxation and may reduce the capacity of conventional electricity; renewables subsidies result in more intermittent wind and solar resources and reduce reliability, etc. Therefore, these policy targets should be studied together. To the best of our knowledge, Reference Ritzenhofen et al. (2014) is the only study that considers all three policy targets together and analyzes the performance of the current markets under high uncertainty. In addition, majority of the existing studies *analyze* the current markets in order to evaluate their performance with respect to reliability, sustainability and price efficiency. However, in our opinion, the correct approach for addressing these problems is a *design* approach starting from the desired policy targets.

We next study the electricity spot market with the loop flow effect. Loop flow effect ties the flow of power in each line to all the other flows in the network. As a result, the production of electricity at each nodes is tied to the production at all other nodes. This way a bilateral contract between a GenCo and demand will affect the possible dispatch of electricity at all other nodes. Therefore, allocating optimal dispatch of electricity, constrained by the electricity transmission network constraints, cannot be achieved by free markets. Currently, there are two approaches to electricity market design with network constraints which use two different sets of prices Ehrenmann and Neuhoff (2009): integrated market design that uses nodal prices (one price for each node in the network); and coordinated market design that uses prices for allocating transmission rights (one price for transmission right between any two nodes) and prices for allocating electricity dispatch (one price for each node in the network). Integrated market design is proposed in the form of physical transmission right (PTR), financial transmission right (FTR) and flow gate right (FGR)(Kristiansen, 2004). In competitive markets with full information, both the integrated market design and the coordinated market design lead to welfare maximizing outcome (See Bohn et al. (1984) & Cardell et al. (1997) for integrated, and Chao and Peck (1996) for coordinated approach). When market power exists (even under full information) the current design of these markets is not efficient Wilson (2008) and the comparison between the performance of the two approaches is not completely known Ehrenmann and Neuhoff (2009). The problem becomes even more challenging when a producer owns multiple electricity generation firms located at different nodes, or when there are multiple producers located at a single node Cardell et al. (1997). In the same way, the comparison between different approaches to coordinated markets is not clear (Joskow and Tirole, 2000).

Finally, we study security of large-scale networks (Chapter 4). The defense of cyber networks, which are typically very large, plays a crucial role in their efficient operation. In these networks, the defender's information grows with the growth in the size of the network. Moreover, the defender faces uncertainty: he has only imperfect observations of the attack events and system state, and is also uncertain about the attacker's strategy. This uncertainty results in a complicated information structure which grows over time. Since the defender is uncertain about attacker's strategy, he should use a robust approach to minimize the worst-case damage to the system among all attacker strategies (minmax approach). Due to these complexities, the design of optimal defense policies is infeasible.

1.2 Contribution of the Thesis

The contribution of this thesis is in the context of two applications: electricity markets and network security.

Our contribution to electricity markets is in two parts. First, we set the loop flow effect (Kirchhoff's laws) aside. We design electricity markets that are reliable, sustainable and price efficient (Chapter 2). Next, we add the loop flow effect and focus on designing electricity spot markets that are price efficient (Chapter 3).

In the first part (Chapter 2), we start by modeling an oligopoly of GenCos which interact through the electricity network for a long horizon and have common uncertainties with respect to the future. The demand for electricity is elastic but non-strategic i.e. demand is a price-taker. The non-strategic demand means that the system operator wants to *auction* the demand among the strategic GenCos. In this model, each GenCo's electricity generation is constrained by its capacity, and the total power flow over each line of the network is constrained by the lines thermal capacity. Next, we add the policy targets to the model. The reliability target is modeled by a constraint requiring that the sum of the GenCos' capacities exceed the planning reserve requirement. The sustainability target is modeled by a constraint requiring that the sum of the carbon emissions is below the carbon cap. We model price efficiency as a desired feature of electricity market design at equilibrium. In addition to achieving price efficiency, we consider the following desired features for the markets at equilibrium (since we only consider features at equilibrium, hereafter we may drop the term[2]):

[2]We only consider these features at equilibrium. In practice, the market clearing process may stop off equilibrium.

F1) social optimality, F2) individual rationality, and F3) budget balanced. Social optimality means that the amount of electricity investment and generation at equilibrium maximizes social welfare. Individual rationality means that the GenCos' utility is not less than their opt out option i.e. the utility they could get by not participating in the market. Budget balance means the price paid to producers is equal to the price charged to the demand. For elastic demand, budget balanced means price should be equal to the marginal utility of the demand[3]. We call a market/auction that possesses features (F1)-(F3) and is price efficient an *efficient market/auction*. Our goal is to design efficient electricity markets considering our model and the policy constraints.

To achieve our market design goal, we develop a framework for designing efficient auctions with constraints. Some of the constraints in our model are convex individual constraints (or simple constraints) such as each GenCo's capacity limit; other constraints are joint constraints (or complex constraints) such as network, reliability and carbon emission constraints. Implementing individual constraints requires a different market design approach than the one required for joint constraints. Therefore, we develop our efficient auction design framework in three steps: (i) designing efficient auctions without constraints; (ii) designing efficient auctions with convex individual constraints; and (iii) designing efficient auctions with homogeneous (additive) joint constraints. In the first step, we propose efficient auctions for both elastic and inelastic demand. In the second step, we show that the proposed markets are also efficient if the auctions are subject to convex individual constraints. Finally, considering the fact that the joint constraints in our model are homogeneous, we introduce a *decomposition technique* that decomposes an auction with homogeneous joint constraints into a sequence of auctions with individual constraints.

Our market design framework has the following implications with respect to electricity policy targets and the debates on implementing them. For each joint constraint we add one market to operate along with the spot electricity market. These markets include capacity market, carbon permit market, and transmission right market. Efficient design of these markets can implement the policy targets. Our results show that price efficiency can be achieved without price caps or offer caps. Also capacity and carbon markets can be used for reliable investment and efficient carbon pricing. Although we do not have a complete answer whether reliability can be implemented by energy-only solutions, but we show that efficient capacity market pays less subsidy for reliability than operation reserve solutions proposed by Hogan et al. (2005).

In the second part (Chapter 3), we study designing efficient electricity spot market when loop flow effect exists (designing networked market). We look at the electricity network as a public good and determine how producers and consumers must use it. Since electricity flow equations

Considering desired features off equilibrium is not the topic of this paper and is a topic for future research.

[3]The inelastic demand does not change its consumption based on the price of electricity and can be charged any price to balance the payment to the producers. This means budget balanced is not a restriction for the inelastic demand.

generally have complex sinusoid form, we consider first order (also known as DC) and second order approximation of these equations. The DC approximation does not include losses over the transmission lines. We show that even though for the DC approximation the set of feasible electricity productions of all the nodes is convex, this set is not convex when considering losses. However, the set of feasible *angles (phases) of the nodes* remains convex even with the losses. Since the angles of the nodes uniquely determine the flow of power in the network, we consider angles as an alternative representation of the productions when losses exist. Moreover, considering angles changes the public good problem into a local public good problem in which the producer/consumer at each node is only affected by the angels of the nodes that are directly connected to it. Such a local public good problem allows for designing scalable mechanisms. We design an efficient (local) public good mechanism that can be used to determine either the productions (in DC approximation) or angles (when losses exist).

Our contribution to the security of large-scale networks is decomposing the system into sub-systems and presenting heuristics for the defense of each sub-system. Each sub-system defense problem possesses a subset of the information in the system. We introduce the notion of *influence graph* that captures the interactions among the network elements. We partition the influence graph by clustering using community detection algorithms. Each region of the network corresponding to each sub-graph is then a sub-system that is controlled by a local controller. We design the information that should be exchanged among local controllers. We formulate a minmax problem from each subsystem and propose a suboptimal solution that is computationally feasible. Even though we do not model the attacker as a strategic decision maker, we compensate (in part) for the lack of this feature by adopting a minmax performance criterion, which leads to a conservative defense approach.

CHAPTER 2

Markets with Complex Policy Constraints: Sustainable, Reliable and Price Efficient Electricity

2.1 Introduction

In this chapter, we study implementation of electricity policy targets by design of electricity markets. We model the policy targets as complex constraints and develop a framework of auctions with constraints for addressing our problem.

Electricity policies in most countries target providing *reliable* and *sustainable* electricity with *efficient price* to customers (IEA, 2016; Ritzenhofen et al., 2014; UNDP, 2015). Reliability refers to meeting the uncertain demand almost at all times (one blackout every ten years), sustainability refers to overcoming the global warming, and efficient price (affordablity) refers to prices close to marginal cost of production. System operators translate reliability target into *planning reserve requirement* that is the total capacity should exceed the maximum possible demand by a reserve margin (%15 to %20). Sustainability target is also translated into *carbon cap* which is a cap for total carbon emission. Planning reserve requirement and carbon cap are *uneconomic constraints* (public constraints) because Generation Companies (GenCos) do not have incentives to meet them unless appropriate taxes/subsidies are in place. Addressing the policy targets for the electricity industry should be consistent with the spirit of the restructured industry i.e. should be based on market competition, incentives and minimum regulations. Moreover, it should consider the physical limits of the electricity grid, mainly electricity transmission line constraints (network constraints).

Before restructuring, the system operator could set the selling price and investment level, and in this way could ensure reliability and price efficiency. After restructuring however, the current electricity markets suffer from scarce resources (capacity underinvestment) and high prices above marginal cost of production; for example in the California 2000 electricity crisis, market imperfections such as market power and *missing money problem*[1] led to very high prices(Joskow

[1]The missing money problem refers to the revenue gap for the investors (investors can not collect their cost of

and Tirole, 2007). Moreover, current mechanisms for reducing carbon emission like carbon tax and carbon markets or renewable feed-in tariffs are inefficient and costly (Alizamir et al., 2016; Borenstein, 2012), specially with larger share of the renewables (Acemoglu et al., 2014, 2016a; Ahlstrom et al., 2015; Baldick, 2012; Hogan, 2015a). There are major debates on and approaches to the correct solutions for these problems: The first is centered around the need for non-market mechanisms like price-cap, offer-cap and market-monitoring so as to achieve efficient prices. The second is about the need for mechanisms with direct capacity incentives, like capacity subsidy or capacity market, so as to achieve the reliability target; this approach is referred to as *energy-and-capacity* solution in contrast to *energy-only* solution. Even without planning reserve requirements, some references argue that capacity mechanisms are required because GenCos' investment will not be socially optimal due to their market power (Garcia and Shen, 2010; Murphy and Smeers, 2005; Novy-Marx, 2007). The third is with regard to carbon tax that is introduced so as to achieve sustainability target and refers to the need for the system operator to set the carbon tax (price policy) or the possibility of finding the efficient carbon tax by carbon markets (quantity policy). On the sustainability topic, there is also an ongoing debate tied to political issues on charging carbon vs. supporting renewables. We briefly discuss this debate in Section (2.7.2).

While the above debates are still ongoing, the electricity industry is under major technological changes, with the most prominent ones being smart grids and renewable technologies [2]. Smart grids introduce demand response and elastic demand to the spot market (Ahlstrom et al., 2015; Joskow and Tirole, 2006). They also alter the definition of reliability from preventing blackouts due to inelastic demand exceeding capacity (Fig. 2.1) to preventing high scarcity prices due to saturated capacity (Fig. 2.2) (Euroelectric, 2010; Henriot and Glachant, 2013; Hogan, 2015b). Renewables discourage investment on electricity in two ways. First, their low marginal cost of production reduces the prices in the spot market and therefore makes it harder for both renewable and conventional utilities to recover their investment costs (Sisternes and Parsons, 2016). Second, renewables' high uncertainty makes them a risky investment (Aïd et al., 2011; Cramton et al., 2013; Henriot and Glachant, 2013; Pritchard et al., 2010; Secomandi and Kekre, 2014). On the other hand, lower carbon emission of renewables can make them an interesting investment for producers if appropriate taxes/subsidies are in place (Acemoglu et al., 2014).

The existing studies of electricity policies are limited to addressing reliability, sustainability and price efficiency separately and in isolation; however these issues are in fact intertwined (Kök

investment) and is due to market imperfections. Investors recover their investment cost when their capacity is fully utilized and the (scarcity) price is beyond their marginal cost of production (Fig. 2.1); therefore, if the demand does not reach full capacity or the price is not allowed to go high enough, scarcity price, which is the market signal for investment, is distorted and investors can not collect their cost of investment; thus, investors reduce their investment.

[2] Global renewable share of the electricity production increases from 21% in 2015 to 29% in 2040. The share of wind and solar from the renewables is increasing from 20% in 2012 to 35% in 2040. Also, some regions like California or Germany will have higher share of the renewables in their electricity generation (IEO, 2016).

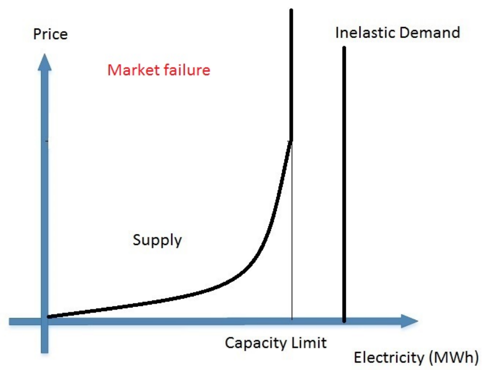


Figure 2.1: Scarcity price: When inelastic demand exceeds capacity, market fails to determine scarcity price.

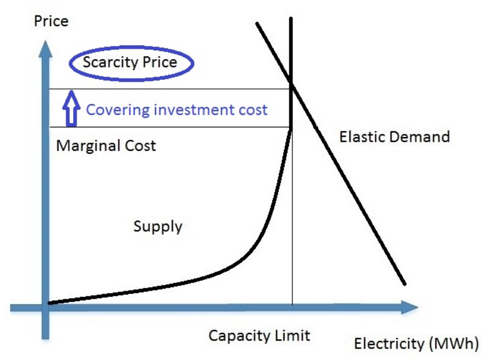


Figure 2.2: Scarcity price: With elastic demand, market determines scarcity price at peak load.

et al., 2016). For example, any changes in the spot market for affordable prices, like price caps, can reduce reliability by reducing investment incentives (Roques and Savva, 2009); reliability subsidies, like capacity investment subsidy, work against carbon taxation so as to reduce the capacity of conventional electricity; renewables subsidies result in more intermittent wind and solar resources and reduce reliability, etc. Therefore, these policy targets should be studied together. To the best of our knowledge, Reference Ritzenhofen et al. (2014) is the only study that considers all three policy targets together and analyzes the performance of the current markets under high uncertainty. In addition, majority of the existing studies *analyze* the current markets in order to evaluate their performance with respect to reliability, sustainability and price efficiency. However, the correct approach for addressing these problems is a *design* approach starting from the desired policy targets.

Our contribution to electricity policy is the following. We consider emerging electricity technological changes and design electricity markets that are reliable, sustainable and price efficient. We start by modeling an oligopoly of GenCos which interact through the electricity network for a long horizon and have uncertainties with respect to the future. The demand for electricity is elastic but non-strategic i.e. demand is a price-taker. The non-strategic demand means that the system operator wants to *auction* the demand among the strategic GenCos. In this model, each GenCo's electricity generation is constrained by its capacity, and the total power flow over each line of the network is constrained by the line thermal capacity[3]. Next, we add the policy targets to the model. The reliability target is modeled by a constraint requiring that the sum of the GenCos' capacities exceed the planning reserve requirement. The sustainability target is modeled by a constraint requiring that the sum of the carbon emissions is below the carbon cap. We model price efficiency as a desired feature of electricity market design at equilibrium. In addition to price efficiency, we consider the following desired features for the markets at equilibrium (since we only consider features at equilibrium, hereafter we may drop the term[4]): F1) social optimality F2) individual rationality and F3) budget balanced. Social optimality means that the amount of electricity investment and generation at equilibrium maximizes social welfare. Individual rationality means that the GenCos' utility is not less than their opt out option i.e. the utility they could get by not participating in the market. Budget balance means the price paid to producers is equal to the price charged to the demand. For elastic demand, budget balanced means price should be equal to the marginal utility of

[3]Network constraints are major limitations in electricity industry. The network has two effects (Downward et al., 2010; Ehrenmann and Neuhoff, 2009). First, the electricity flow in each line is limited by the maximum capacity of the line that is determined by thermal (and other physical) limits. This is the same network effect as in transportation or communication networks. Second, the entire injection profile in the network should satisfy KVL and KCL constraints. This is a unique feature of electricity networks; it does not exist in other networks. Corresponding voltage and phase laws create negative externality effects. Network constraints cause externalities and result in locational market power and locational prices (Ehrenmann and Neuhoff, 2009). In this paper we only study the first effect i.e. lines' thermal capacity.

[4]We only consider these features at equilibrium. In practice, the market clearing process may stop off equilibrium. Considering desired features off equilibrium is not the topic of this paper and is a topic for future research.

the demand[5]. We call a market/auction that possesses features (F1)-(F3) and is price efficient an *efficient market/auction*. Our goal is to design efficient electricity markets considering our model and the policy constraints.

To achieve our market design goal, we develop a framework for *designing efficient auctions with constraints*. Some of the constraints in our model are *convex individual constraints* (or simple constraints) such as each GenCo's capacity limit; other constraints are *joint constraints* (or complicating constraints) such as network, reliability and carbon emission constraints. Implementing individual constraints requires a different market design approach than the one required for joint constraints. Therefore, we develop our efficient auction design framework in three steps: (i) designing efficient auctions without constraints; (ii) designing efficient auctions with convex individual constraints; and (iii) designing efficient auctions with homogeneous (additive) joint constraints. In the first step, we propose efficient auctions for both elastic and inelastic demand. In the second step, we show that the proposed markets are also efficient if the auctions are subject to convex individual constraints. Finally, considering the fact that the joint constraints in our model are homogeneous, we introduce a *decomposition technique* that decomposes an auction with homogeneous joint constraints into a sequence of auctions with individual constraints.

Our market design framework has the following implications with respect to electricity policy targets and the debates on implementing them. First, our efficient design for auctions without constraints shows that without network, reliability and carbon constraints, efficient prices (along with features F1-F3) are achievable by the correct design of the spot market and without using offer cap, price cap, market monitoring or other sorts of market interventions. Second, our results contribute to the debate on energy-and-capacity vs. energy-only solutions both with and without planning reserve requirements. Without planning reserve requirements, our results for auctions with individual constraints contradict the argument of References Garcia and Shen (2010) and Murphy and Smeers (2005) by showing that if our proposed spot market is used (instead of Cournot markets studied in these references), the energy-only mechanism is capable of implementing the socially optimal investment. When planning reserve requirements exist, our framework can be used for implementing reliable investment in two ways: First, since planning reserve requirement is a homogeneous joint constraint, we can use our decomposition technique to design efficient capacity and spot markets for implementing it. We obtain an energy-and-capacity solution where the price in the capacity market is the efficient subsidy for reliability target; this subsidy is positive only if the reliability constraints are binding. Second, the planning reserve requirement can also be implemented by increasing the demand in the spot market above the real demand (Hogan et al., 2005); the extra demand is put in operation reserve. We obtain an energy-only solution and our

[5] The inelastic demand does not change its consumption based on the price of electricity and can be charged any price to balance the payment to the producers. This means budget balanced is not a restriction for the inelastic demand.

results for auctions with individual constraints can be used for determining demand profiles that achieve the reliability target with minimum subsidy required in an energy-only approach We show that this minimum payment can be greater than the subsidy paid in an efficient capacity-and-energy approach with efficient capacity markets[6]. The implication of our results for implementing carbon constraints is the following.

By defining *carbon permit* as the level of carbon each producer is allowed to emit, the carbon constraints can be translated from joint constraints on productions into homogeneous joint constraints on carbon permits which, by decomposition technique, can be implemented with efficient carbon permit markets (or simply carbon market). The price of carbon permit found in the carbon market is the efficient level of carbon tax. The carbon tax is positive only if the carbon constraint is binding. With respect to the debate on carbon market vs. carbon tax, this result shows that the system operator can determine efficient carbon tax by the carbon permit market without knowing all the private information of the GenCos. Of course, this comes at a higher regulatory cost of running the carbon market. Our results also contribute to the debate on charging carbon vs. supporting renewables; we briefly discuss this issue in Section (2.7.2). Finally, the lines' thermal capacity constraint can also be implemented in the same way as the carbon emission constraint: By defining *transmission right* as the amount of electricity power a GenCo is allowed to send over a certain line in a certain direction, the lines' thermal capacity can be modeled by homogeneous joint constraints on the GenCos' transmission rights. This allows us to use a decomposition technique for designing efficient transmission right markets that allocate the transmission rights by determining the efficient transmission tax[7]. This tax is positive only if the line is congested in the corresponding direction. We note that the proposal for transmission right exists in the literature (Chao and Peck, 1996; Ehrenmann and Neuhoff, 2009; Hogan, 1992), but its efficient market implementation has been an open problem.

Central to our theoretical results is the assumption of symmetric information among GenCos and the use of Nash equilibrium as a solution concept. This assumption has both theoretical and practical reasons. Theoretically, eliminating information asymmetry and its corresponding information rent allows us to develop a basic understanding of our problems and rigorous analytical results. This assumption is made in majority of the literature (for example see Allaz and Vila (1993), Ehrenmann and Neuhoff (2009))[8]. A number of papers argue that symmetric information

[6]The other practical challenge of this energy-only solution is the calculation of the demand profile. In real world, calculation of the efficient demand profile requires a lot of information and complex computations by the system operator; moreover, the profile should be continuously updated based on the changes in the industry.

[7]This market for transmission permit is for optimal dispatch and is different from the transmission right market that is designed to hedge GenCos against uncertainty. Moreover, transmission permit is the first step in understanding the network effect in electricity. In practice, the loop flow effect of the electricity network due to Kirchhoff's laws makes the problem much more complicated.

[8]Relaxing the assumption of symmetric information comes at the cost of suboptimal outcomes. See the literature of

is a realistic assumption for electricity markets. For example, Reference Wang and Zender (2002) argues that symmetric information of the bidders (GenCos) which is also superior to the information of the auctioneer (system operator) is a common rationale and a reliable assumption for oligopolistic electricity markets. This is because GenCos monitor each others' technologies and capacity, but the ISO does not have access to such information. Symmetric information does not exclude uncertainty. In fact, uncertainties in the electricity industry can be modeled by considering *common/symmetric uncertainty* among the GenCos. In our model, we consider symmetric uncertainty and show that for risk-neutral GenCos who want to maximize their expected profit over a time horizon, the results obtained without uncertainty do not change.

Moreover, GenCos in electricity industry have long-term strategic interaction. Long-term strategic interaction allows GenCos to cooperate/collude their strategies over time and form a monopolistic supply. This makes the design of dynamic electricity auctions difficult. Our efficient dynamic auctions forbid the possibility of collusion in finite horizon interactions. For infinite horizon interactions, collusion is avoided if the GenCos only use pay-off relevant information in their strategies. References Butler and Neuhoff (2008); Novy-Marx (2007) and Garcia and Shen (2010) argue that pay-off relevant strategies are a valid assumption for electricity industry.

2.1.1 Literature Review

We investigate the literature related to our work by first studying the literature on price efficient auctions. Next, we investigate the literature on investment without reliability targets (without planning reserve requirements) followed by the literature on investment with reliability targets. Sustainability (carbon emission constraint) and network lines capacity constraints are studied afterwards. We finish this section by studying the literature on symmetric uncertainty in long-term interactions among agents.

Price efficient auctions for electricity are static auctions of divisible goods. Static auctions can be categorized into uniform and discriminatory price auctions. Supply function (SF), Cournot, first and second price auctions are examples of uniform price auctions. In SF auctions GenCos bid their cost curve and the market is cleared by dispatching minimum cost first (Green and Newbery, 1992; Klemperer and Meyer, 1989). In practice, SF auctions are used for electricity spot markets where the same bids are used to clear multiple possible demands due to demand uncertainty or demand variation during the day (Anderson and Hu, 2008; Downward et al., 2010; Elmaghraby and Oren, 1999). In general, SF auctions have a range of equilibria and are not efficient (Klemperer and Meyer, 1989). The inefficiency of SF auctions has been studied further within the context of polynomial supply functions (Baldick and Hogan, 2004), oligopoly models with capacity constraints

Bayesian implementation and the more recent literature of robust auction design for more discussion.

(Baldick and Hogan, 2006; Baldick et al., 2001), with pivotal bidder (Genc and Reynolds, 2011), or with limited number of price bids (Wilson, 1979), or with demand uncertainty (Holmberg and Wolak, 2015); it is also studied within the context of duopoly models with finite piece-wise linear cost function (Anderson, 2013). A review of supply function auctions literature is provided in (Holmberg and Newbery, 2010). Cournot auctions are used to approximate SF electricity auctions (Downward et al., 2010); they also do not implement the socially optimal allocation (Anderson and Hu, 2008; Elmaghraby and Oren, 1999). First and second price auctions have been used for capacity markets (Salant and Stoddard, 2008). While these auctions implement the socially optimal allocation in Nash equilibria, they are not price efficient. Discriminatory price auctions can also achieve social optimality but are not price efficient; auctions like Vickery-Clarke-Groves (VCG) and dAspremont-Gerard-Varet (AGV) achieve social optimality by paying each seller an *individual price* that aligns his incentive with the social welfare; however, VCG and AGV auctions can not be budget balanced and individually rational at the same time. In electricity markets, uniform price auctions are preferred because they provide a unique and clear signal for investment (Fabra et al., 2006). We are not aware of an efficient (price efficient, socially optimal, individually rational, and budget balanced) auction of a divisible good in the literature. Our proposed efficient auctions use uniform price at equilibrium to achieve price efficiency and budget balance, and discriminatory prices off-equilibrium to align individuals' incentives with the social welfare. Such an approach is proposed in the British restructuring of the electricity markets (FERC, 2012; Hogan, 2016). Our mechanisms are inspired by Hurwicz (1979) but also consider price efficiency and elastic demand.

Investment without a reliability target (planning reserve requirement) has been studied in the literature of capacity-trade competition where firms first expand their capacity and later use that capacity to compete in the trade market[9]. Capacity-trade competition has been studied in the literature by analyzing specific forms of trade markets including Bertrand (capacity-price competition) (Acemoglu et al., 2009; Davidson and Deneckere, 1986; de Frutos and Fabra, 2011), Cournot (capacity-quantity competition) (Garcia and Shen, 2010; Murphy and Smeers, 2005; Perakis and Sun, 2014; Roques and Savva, 2009) , VCG (Bergemann and Välimäki, 2002) and competitive markets (Dixit, 1991). In the case of Bertrand and Cournot, the market equilibrium is not socially optimal. In the case of VCG and competitive markets, the outcome is socially optimal; but, VCG is not budget-balanced and perfect competition is a strong assumption for electricity oligopolistic markets. Our spot market proposal, in contrast to the markets proposed so far in the literature, achieves the desired features (F1)-(F3) and is price efficient.

Implementing reliability targets for electricity has been studied by considering sources of

[9]In oligopoly markets, increasing investment has two opposite effects resulting in inefficiency: on one hand it can increase the investor's share of the trade market, but on the other hand it can introduce more competition and lower the prices.

underinvestment that are market power and market imperfections leading to missing money problem. Market imperfections are further categorized into demand inelasticity, uncertainty, and network effects/constraints. If the inelastic demand exceeds maximum capacity it results in market failure where supply and demand mismatch without any scarcity price to signal correct investment (Fig. 2.1) (Cramton and Stoft, 2005; Cramton et al., 2013). Uncertainties result in reliability requirements that are not valued by the GenCos as much as they are by the ISO (Sun et al., 2010); if GenCos are risk averse, they have incentives to even further reduce the investment level (Aïd et al., 2011; Pritchard et al., 2010; Secomandi and Kekre, 2014). The network constraints induce externalities such that GenCo's may not be able to use their capacity if the lines to the consumers are congested by others (Borenstein, 2002; Borenstein et al., 2002; Joskow and Tirole, 2007). The methods proposed to address the reliability constraints can be categorized into: (i) correcting the spot (energy) markets e.g. using new market processes, lifting price caps and price-dependent market intervention techniques (Joskow, 2006); (ii) using supportive energy remuneration mechanisms (ERM), e.g. operation reserve markets (Hogan et al., 2013), forward generation contracts (Allaz and Vila, 1993) and mandatory load hedging (MLH) (Hogan et al., 2005); and (iii) using supportive capacity remuneration mechanisms (CRM) e.g. forward capacity obligation with capacity markets/bilateral trades (ICAP), capacity payments, strategic reserve and reliability options (Bidwell, 2005; Cramton and Stoft, 2005; Cramton et al., 2013; Hogan et al., 2005). Supportive mechanisms should be designed in a way that their effect automatically diminishes as the spot market imperfections are resolved (CAISO, 2014; Henriot and Glachant, 2013). While none of the existing solutions to the underinvestment problem is a complete market-based response, we propose two such solutions for reliability: (i) an efficient energy-only solution with operation reserve using efficient auctions with individual constraints; and (ii) an efficient energy-and-capacity solution with capacity markets using efficient auctions with joint homogeneous constraints.

Practices for charging carbon emission have been in the form of carbon emission tax (Acemoglu et al., 2016b) or carbon cap-and-trade market (Borenstein et al., 2015). With a fully-informed system operator, they lead to the same outcome (Weitzman, 1974); otherwise, only an efficient cap-and-trade market can find the efficient tax (Keohane, 2009; Metcalf, 2008). Since directly charging carbon faces political resistance, indirect policies for supporting renewables have been considered (Borenstein, 2012)[10]. These supportive policies include forward contracts in terms of feed-in tariffs (Alizamir et al., 2016; Butler and Neuhoff, 2008; Couture and Gagnon, 2010), generation subsidy for renewables or feed-in premiums (Hobbs, 2015) and subsidy on capacity

[10]There are also other non-environmental arguments for supporting renewables including energy security, green jobs, non-appropriate intellectual property, and lowering the cost of fossil fuels (see (Borenstein, 2012)). Also renewables are an economic solution for off-grid consumers including islands and rural areas. There are more than 50 thousand islands on earth with a total area of over one sixth of the global land area, and more than 740 million inhabitants (Kuang et al., 2016).

expansion (Sisternes and Parsons, 2016). In addition, a recent line of papers have argued alternative approaches for making renewables an attractive investment; Reference Aflaki and Netessine (2016) proposes approaches to reduce the intermittency of renewables; Reference Kök et al. (2016) argues for flat pricing of electricity. Current policies for charging carbon and supporting renewables are not efficient or practical with larger share of the renewables (Alizamir et al., 2016; Borenstein, 2012). Calculation and continuous adjustment of the carbon tax is a difficult task. On the other hand, current carbon markets are not efficient as GenCos can use their market power to pay a price for carbon that is lower than the efficient tax. Also, renewable supportive mechanisms that require a considerable financial support for renewables, result in negative spot prices, incentivize investment only indirectly and inefficiently, are price insensitive resulting in inelastic supply, and can be in conflict with reliability targets because they decrease the investment on reliable conventional technologies (Acemoglu et al., 2014, 2016a; Ahlstrom et al., 2015; Baldick, 2012; Borenstein, 2012; Hogan, 2015a). In contrast to the current supportive policies, renewables should be introduced to the regulations of the competitive markets (Ahlstrom et al., 2015; Henriot and Glachant, 2013; Sisternes and Parsons, 2016). Such markets, if designed correctly, can automatically determine and update the efficient amount of financial support for renewables. Designing efficient markets for both charging carbon and supporting renewables is an open problem (Ahlstrom et al., 2015; Henriot and Glachant, 2013; Sisternes and Parsons, 2016). Our model and results for auctions with homogeneous joint constraints along with the proposed decomposition technique can be used for designing such markets.

Electricity network constraints result in local market power and inefficiency. This inefficiency has been studied for Cournot (Downward et al., 2010) and supply function (Wilson, 2008) auctions. Reference Blazquez de Paz (2014) studies the inefficiency due to the network constraints when financial transmission rights exist. Even for a stylized model of a bipartite network with infinite line capacities, Reference Bimpikis et al. (2014) shows the inefficiency of the Cournot competition. Reference Chao and Peck (1996) identifies line capacity constraints in electricity spot markets as homogeneous joint constraints and proposes using markets for allocating *transmission rights* between each two nodes, but the design of such market is left as an open question. References Ehrenmann and Neuhoff (2009) and Hogan (1992) study transmission rights in comparison to nodal pricing, and argue that under competitive markets and Cournot auctions, both the transmission right approach and the nodal pricing approach lead to socially optimal outcome; however, with existence of market power none of these approaches is efficient and their comparison is not clear. Our model for homogeneous joint constraints along with the proposed decomposition technique leads to the design of efficient transmission right markets.

Symmetric uncertainty has been studied within the context of capacity-trade competition. Reference (Bushnell and Ishii, 2007) provides an energy-only model (a model of only spot markets) for capacity expansion under uncertainty and by finding a Markov perfect equilibrium using simulations, it shows that the underinvestment problem exists in this model. Reference (Garcia and Shen, 2010) uses a highly stylized model with Cournot spot markets and provides the same results analytically. GenCos' risk aversion facing uncertainty is studied in Reference Secomandi and Kekre (2014) which shows the advantage of forward contracts over real-time contracts for the dispatch of the predicted demand. Reference Aïd et al. (2011) shows that with highly risk averse consumers vertical integration is superior to risk hedging. Reference Pritchard et al. (2010) develops a programming method for determining efficient dispatch and price of electricity considering uncertainty of renewables. We consider risk-neutral GenCos in this paper. Our model of uncertainty and long-term interaction of the producers is similar to References Bushnell and Ishii (2007); Garcia and Shen (2010), but in contrast to these papers, we design markets that are efficient in this setup.

The rest of the paper is organized as follows. In Section (3.2) we present our model. In Section (2.3) we propose efficient auctions for allocation of an elastic or inelastic divisible good among agents. We study efficient auctions for the allocation of a divisible good under individual constraints in Section (2.4). In Section (2.5), we study efficient auctions for the allocation of a divisible good under homogeneous joint constraints; to this end, we present a technique for decomposing the allocation problem into a sequence of allocations with individual constraints; each allocation with individual constraints is implemented with an efficient auction. We discuss long-term interaction of GenCos with symmetric uncertainty in Section (2.6) by extending the model and results of Section (2.5)[11]. In this Section, we use the results of Sections (2.3)-(2.6) to design efficient auctions of electricity policy targets in Section (2.7). In Section (2.8) we conclude the paper and present future directions.

2.2 Model

We first model the electricity network and then the policy targets. Other desired features for electricity markets are subsequently added to the model. Remarks on comparing the model with the real world and extensions on the model are discussed at the end of this section [12].

Consider a T-period (T-year) economy, $T = \{1,2,...,T\}$, consisting of $N = \{1,2,...,N\}$ strategic GenCos ($N \geq 3$ for technical issues), a non-strategic elastic demand with discount factor β and an independent system operator (ISO). GenCos have symmetric information with symmetric uncertainty. The random variable $\mathbf{w}_t \in W_t$ models the uncertainty about the future production costs,

[11] In Appendix D, we present the same extensions for the results of Section (2.4).

[12] In this section and consequent sections, bold parameters refer to random variables. The realizations of the random variables are not bold. Capital letters refer to sets. Also for convenience we use the notation $a_{x:y}$ to refer to $\{a_x, a_{x+1}, ..., a_{y-1}, a_y\}$. For variables $a_1, a_2, ..., a_N$ if their range N is known, then a without subscript refers to $a_{1:N}$. For example $\{e_{1,t}, ..., e_{N,t}\}$ is referred to by e_t.

expansion costs, carbon emission factor and demand utility as well as regulatory uncertainty about reliability target and carbon emission cap at time t. The distribution of the uncertainties, $f(\mathbf{w}_{1:T})$, is common knowledge among GenCos and the ISO. Let $e_{n,t}$ denote GenCo n's electricity production at time t. GenCo n's production cost at time t is $C^e_{n,t}(e_{n,t},\mathbf{w}_t)$ where $\forall w_t \in W_t$, $C^e_{n,t}(e_{n,t},w_t)$ is strictly convex and increasing and $C^e_{n,t}(0,w_t) = 0$. The demand's utility at time t is denoted by $U_t(d_t,\mathbf{w}_t)$, where $\forall w_t \in W_t$, $U_t(d_t,w_t) = 0$ and $U_t(d_t,w_t)$ is strictly concave and increasing. Let $\Delta x_{n,t}$ denote GenCo n's capacity expansion at time t. GenCo n's expansion cost at time t is limited by $\overline{\Delta x}_{n,t}(\mathbf{w}_{t-1})$ determined by its physical and budgetary limits. GenCo n's expansion cost at time t is $C^x_{n,t}(\Delta x_{n,t},\mathbf{w}_{t-1})$. For every $\mathbf{w}_{t-1} \in \mathbf{w}_{t-1}$, $C^x_{n,t}(\Delta x_{n,t},\mathbf{w}_{t-1})$ is strictly convex and increasing with respect to $\Delta x_{n,t}$ with $C^x_{n,t}(0,\mathbf{w}_{t-1}) = 0$. The GenCos and the demand connected via an electricity network. The network consists of $V = \{1,2,...,V\}$ buses/nodes (vertices) and L lines (edges) connecting them. The elastic non-strategic demand is located in node 1. Each GenCo $n \in N$ is located in one of the other $V-1$ nodes which is denoted by v_n. The set of producers in node $v \in V$ is denoted by N_v. The line between two nodes $u,v \in V$ has thermal capacity $\bar{I}_{uv}$ of sending power flow from u to v. The lines' thermal capacities are symmetric i.e. $\bar{I}_{uv} = \bar{I}_{vu}$. We denote by $e_{n,uv}$ GenCo n's power flow from node u to node v; negative values of $e_{n,uv}$ represent a positive power flow from v to u. The total power flow from u to v at time t is $I_{uv,t} = \sum_{n\in N} e_{n,uv,t}$. This total power flow is constrained by the lines' thermal capacities (first network constraint),

$$-\bar{I}_{vu} \leq \sum_{n\in N} e_{n,vu,t} \leq \bar{I}_{vu}. \tag{2.1}$$

Moreover, GenCo n's total outflow of power from node $v \in V$ at time $t \in T$ is the sum of GenCo n's power flow from v to all other nodes $u \in V$ at that time, $\sum_{u\in V} e_{n,vu,t}$; GenCo n's outflow from the node v_n where it is located at should be equal to GenCo n's generation at time t (second network constraint),

$$\sum_{u\in V} e_{n,v_n u,t} = e_{n,t}, \quad \forall t \in T. \tag{2.2}$$

GenCo n's outflow of power from any node other than its production node, v_n, or the demand node, node 1, should be zero (third network constraint).

$$\sum_{u\in V} e_{n,vu,t} = 0, \quad v \in V \backslash \{1, v_n\} \quad \forall t \in T. \tag{2.3}$$

The second and the third network constraints imply that the inflow of producer n to the demand node, node 1, is equal to its total production, $\sum_{u\in V} e_{n,1u,t} = -e_{n,t}$; this is because all the produced electricity is consumed at node 1.

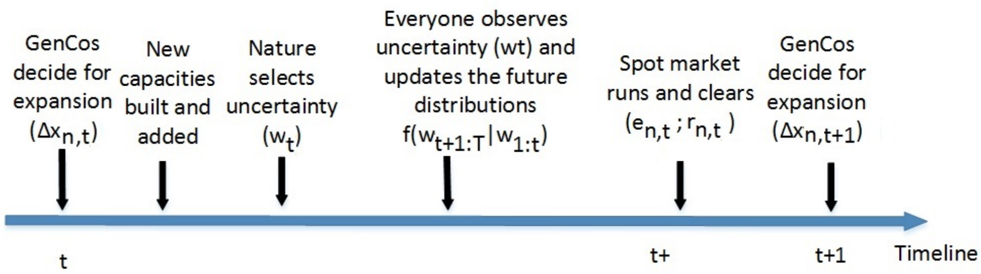


Figure 2.3: Timeline of events in the electricity industry during period t to $t+1$.

Initially, GenCos have capacity $x_{n,0}, n \in N$. At the beginning of each year $t \in T$, the GenCos simultaneously and independently decide on how much to expand their existing capacities. The new capacities are built and added to the system so that the total GenCo n's capacity for generating electricity at time t is $x_{n,t} = x_{n,0} + \sum_{\tau=1,2,\ldots,t} \Delta x_{n,\tau}$[13]. Next, nature selects the realization of the uncertainty for year t, w_t. All GenCos and the ISO observe w_t and update their belief about future uncertainties $f(\mathbf{w}_{t+1:T}|w_{1:t})$. After the realization of the uncertainty, demand at time t^+ with utility $U_t(d_t, w_t)$, is cleared. The timeline of events during year t is shown in Fig. 2.3. After the final year, at time $T+1$, each GenCo $n \in N$ sells its entire capacity with price $\eta(\mathbf{w}_t)$, gains a value of $V_c(x_{n,T+1}, \mathbf{w}_{T+1}) = \eta(\mathbf{w}_{T+1}) \times x_{n,T+1}$ and leaves the economy. This is the model for the industry.

Next, we model the system operator's policy targets i.e. reliability, sustainability and price efficiency. The reliability target is in the form of the following planning reserve requirements:

$$\sum_{n\in N} x_{n,t} \geq \underline{x}_t(\mathbf{w}_{t-1}), \quad \forall t \in T. \tag{2.4}$$

This means that at each time t, there is a minimum capacity requirement; this capacity requirement depends on the realization of uncertainty up to time $t-1$ and is denoted by $\underline{x}_t(\mathbf{w}_{t-1})$. Assuming that constructing the minimum required capacity is feasible, the expansion limits of the GenCos should satisfy

$$\underline{x}_t(\mathbf{w}_{t-1}) \leq \sum_{n\in N} x_{n,t} \leq \sum_{n\in N} x_{n,0} + \sum_{\tau=1,2,\ldots,t} \sum_{n\in N} \bar{x}_{n,\tau}(\mathbf{w}_\tau) \quad \forall t \in T, \mathbf{w}_{1:t} \in W_{1:t}. \tag{2.5}$$

We model the system operator's sustainability target as follows. Consider $\alpha_{n,t}(\mathbf{w}_{t-1})$ to be the carbon factor for the technology of producer n used for producing electricity at time t. Then at time

[13]While we assume that the new capacities are immediately built and available at the same year, in practice it takes a number of years to build and add them to the system. We discuss this issue further in Section (2.8).

t, the sustainability target restricts carbon emission by carbon cap $\overline{B}_t(\mathbf{w}_{t-1})$ as follows

$$\sum_{n\in N} \alpha_{n,t}(\mathbf{w}_{t-1})e_{n,t} \leq \overline{B}_t(\mathbf{w}_{t-1}) \tag{2.6}$$

Price efficiency requires that a uniform price is paid to all GenCos and this price is equal to the marginal cost of production for those producers that their production positive and is not saturated from above by their capacity. Assume $p_{n,t}$ to be is the price paid to GenCo n at time t. Then, price efficiency requires that $\forall n \in N, t \in T$

$$p_{n,t} = p_t \tag{2.7}$$

$$p_{n,t} = C'_{n,t}(e_{n,t}) \quad \text{if } 0 < e_{n,t} < x_{n,t} \tag{2.8}$$

Budget balance, social welfare maximization, and individual rationality are the other desired features of electricity markets. Budget balance is defined when the demand is elastic and requires the sum of money paid to the producers to be equal to the money collected from the demand. Since $\frac{\partial U_t(d_t,\mathbf{w}_t)}{\partial d_t}$ is the price per unit of electricity that the elastic demand is paying for consuming d units of electricity, budget balance means that for each $w_t \in W_t$,

$$\sum_{n\in N} p_{n,t}e_{n,t} = \frac{\partial U_t(\sum_{n\in N} e_{n,t}, w_t)}{\partial(\sum_{n\in N} e_{n,t})} \times \sum_{n\in N} e_{n,t}, \forall t \in T \tag{2.9}$$

which for the case of $\sum_{n\in N} \hat{e}_{n,t} \neq 0$ means

$$p_t = \frac{\partial U_t(\sum_{n\in N} e_{n,t}, w_t)}{\partial(\sum_{n\in N} e_{n,t})}. \tag{2.10}$$

In other words, the price should follow the demand-price curve. f Inelastic demand pays any amount for electricity consumption; therefore budget balance can be achieved by charging the demand the total money paid to the GenCos.

Social welfare maximization is determined by appropriate choice of investment and production plans that are functions of the system's history of the system. At year t, we denote the history for expansion by h_t and the history for production by h_{t^+} and their spaces by H_t, H_{t^+}. Define $E_{1:t}(x_0, \Delta x_{1:t})$ to the space of all possible electricity productions based on the initial capacity x_0 and capacity expansions Δx_t; also define $\Delta X_{1:t}(\overline{x}_{1:t})$ as the space of all possible expansions based on the expansion limits $\overline{x}_{1:t}$. Then,

$$h_t := \{x_0, e_{1:t-1}, \Delta x_{1:t-1}, w_{1:t-1}\} \in H_t := \big(x_0, E_{1:t-1}(\Delta x_{1:t-1}), \Delta x_{1:t-1}(\overline{x}_{1:t-1}), W_{1:t-1}\big) \tag{2.11}$$

$$h_{t^+} := \{x_0, e_{1:t-1}, \Delta x_{1:t}, w_{1:t}\} \in H_{t^+} := \big(x_0, E_{1:t-1}\big(\Delta x_{1:t-1}\big)\big), \Delta x_{1:t}(\overline{x}_{1:t}), W_{1:t}\big) \tag{2.12}$$

GenCo n's investment plan/strategy, $n \in N$, at year t is denoted by $\sigma_{n,t} : H_t \to [0, \overline{\Delta x}_{n,t}(\mathbf{w}_{t-1})]$ and its production plan/strategy at year t is denoted by $\gamma_{n,t} : H_{t^+} \to [0, x_{n,t}]$. We denote the space of all such strategies by $\Sigma_{n,t}$ and $\Gamma_{n,t}$, $t = 1, 2, ..., T$ respectively and define

$$\gamma := \{\gamma_{n,t} : n \in N, t \in T\} \tag{2.13}$$

$$\sigma := \{\sigma_{n,t} : n \in N, t \in T\} \tag{2.14}$$

The investment and production strategies that maximize the expected social welfare are the solution of the following problem:

$$\max_{\sigma_{n,t} \in \Sigma_{n,t}, \gamma_{n,t} \in \Gamma_{n,t} : n \in N, t \in T} E_{\mathbf{w}_{1:T}} \left[\sum_{t \in T} \beta^t U_t \left(\sum_{n \in N} e_{n,t}, \mathbf{w}_t \right) - \right. \qquad \text{(Socially Optimal)}$$

$$\left. \sum_{t \in T} \sum_{n \in N} \beta^t \left\{ C^x_{n,t}(\Delta x_{n,t}, \mathbf{w}_{t-1}) + C^e_{n,t}(e_{n,t}, \mathbf{w}_t) \right\} + \beta^{T+1} \sum_{n \in N} \eta(\mathbf{w}_T) x_{n,T} \right] \tag{2.15}$$

$s.t. \quad \forall t \in T, n \in N, \forall u, v \in V, w_{1:T} \in W_{1:T}$

$$0 \le e_{n,t} \le x_{n,0} + \sum_{\tau = 1,...,t} \Delta x_{n,\tau} \tag{2.16}$$

$$0 \le \Delta x_{n,t} \le \overline{\Delta x}_{n,t}(\mathbf{w}_{t-1}) \tag{2.17}$$

$$\sum_{n \in N} x_{n,t} \ge \underline{x_t}(\mathbf{w}_{t-1}) \tag{2.18}$$

$$\sum_{n \in N} \alpha_{n,t}(w_{t-1}) e_{n,t} \le \overline{B}_t(\mathbf{w}_{t-1}) \tag{2.19}$$

$$-\bar{I}_{vu} \le \sum_{n \in N} e_{n,vu,t} \le \bar{I}_{vu} \tag{2.20}$$

$$\sum_{u \in V} e_{n,v_n u,t} = e_{n,t} \tag{2.21}$$

$$\sum_{u \in V} e_{n,vu,t} = 0, \quad v \in V \backslash \{1, v_n\} \tag{2.22}$$

Finally, individual rationality requires that the expected future utility of each GenCo $n \in N$ at each time $t \in T$ are greater than its outside option, that is the (expected) future utility each GenCo receives by not participating in the production and expansion process. Based on the cost functions, we assume the outside option for not producing or expanding to be zero. If the payment that GenCo n receives at time t is denoted by $r_{n,t}$, then individual rationality at times $t \in T$ and $t^+ \in T$ are expressed by the requirements

$$E_{\mathbf{w}_{t:T}} \left[\sum_{\tau = t, t+1, ..., T} \beta^\tau \left(-C^e_n(e_{n,\tau}, \mathbf{w}_\tau) - C^x_n(\Delta x_{n,\tau}, \mathbf{w}_\tau) + r_{n,\tau} \right) \middle| h_t \right] \ge 0 \quad \forall h_t \in H_t, \tag{2.23}$$

and

$$E_{\mathbf{w}_{t+1:T}}\left[\beta^t\left(-C_n^e(e_{n,t},\mathbf{w}_\tau)+r_{n,t}\right)+\sum_{\tau=t+1,t+2,\ldots,T}\beta^\tau\left(-C_n^e(e_{n,\tau},\mathbf{w}_\tau)-C_n^x(\Delta x_{n,\tau},\mathbf{w}_\tau)\right.\right.$$
$$\left.\left.+r_{n,\tau}\right)\middle| h_{t^+}\right]\geq 0 \quad \forall h_{t^+}\in H_{t^+}, \qquad (2.24)$$

respectively.

Remark 2.2.1 (EXTENSIONS OF THE MODEL)**.** *The model presented here is not a representation of all the details of real-world electricity markets; it can be extended in the following five major directions so as to become closer to reality by taking into account modeling cost/utility functions, capacity expansion, network constraints, asymmetric information, and transmission expansion planning. We briefly discuss these directions below. We further discuss their impact on the design of electricity markets in Section (2.8).*

We first revisit our assumptions for costs and utilities. Although we assumed costs of production and expansion are convex, in real world both production and expansion can have economy of scale and therefore concave Cadwalader et al. (2010). Also, although we assumed the costs are zero for zero production and expansion, in reality there exist start up costs or fixed costs.

The second direction is capacity expansion. While we assume that the new capacities are immediately built and available at the same year, in practice it takes a number of years to build and add them to the system. Also, our model does not consider depreciation of the capacity. The above modeling assumptions are made for simplicity of presentation; the results of this section hold also when there exists a fixed constant rate or an uncertain reliable rate of capacity depreciation. In addition, whereas in practice companies review their expansion decisions each year, there are numerous spot markets during the year (one in almost every 10 *minutes). Therefore, the spot market in this model should be interpreted as an abstract market model that captures the aggregate effect of all the spot markets during a year.*

The next expansion direction of the model is the network. We have assumed there is a single demand at node 1*. In reality demand is distributed over the network. Furthermore, we have assumed each GenCo owns only one electricity production utility located in a certain node. In reality each GenCo owns multiple utilities in different locations. Considering different expansion locations allows GenCos to decide where they want to expand their capacity. More importantly, we have not modeled the loop flow effect due to Kirchhoff's electricity laws. The loop flow effect is what distinguishes electricity networks from other networks (e.g. transportation, communication, etc.). In this paper our goal is to understand the effect of constraints (2.16) - (2.22) on the strategic GenCos' production and capacity expansion plans; for this reason, we did not consider the loop flow effect due to Kirchhoff's laws.*

The fourth major extension to the model is to consider the information asymmetry among the GenCos. We have assumed information is symmetric, but in reality, the GenCos have private information on their costs, technology and budgets.

Fifth, we have not modeled transmission expansion planning. We can extend the model by adding transmission companies (TransCos) and consider their expansion decisions.

In addition to the above four directions, we can consider the following extensions to the model. We presented a model with elastic demand in the spot market. The demand could also be inelastic. We discuss both elastic and inelastic spot markets in Section (2.3). Also, there are other constraints that could be added to the model. For example, ramping constraints or $N-1$ *power flow constraints for short term reliability can be added to the model.*

In order to understand how to design efficient auctions that implement the social choice function of Problem (Socially Optimal), we study a sequence of problems that are increasing in complexity in terms of the constraints they include. These problems deal with the optimization of the the social welfare function (i) without any constraints (Section (2.3)), (ii) with individual constraints of Eqs. (2.16),(2.17) and (2.21)-(2.22) (Section (2.4)), (iii) with joint constraints of Eqs. (2.18)-(2.20) (Section (2.5)), and (iv) with long-term interaction of GenCos and symmetric uncertainty (Section (2.6)).

2.3 Efficient Auction of a Divisible Good

We study the design of efficient auctions for divisible goods without any constraints. This problem serves as the fundamental building block of our development. We consider two separate cases of price elastic and price inelastic divisible goods. We present auction design for electricity spot markets, but the model and the design are general enough so that they can be extended to situations where the mechanism designer is a social welfare maximizer and wants to allocate a divisible good to an oligopoly of agents; for example, the proposed efficient auction of inelastic divisible good will be used in the next sections to design capacity market, carbon permit market and transmission right market.

Inspired by (Hurwicz, 1979), the design is inspired based on the following. (1) The price paid to each strategic producer is independent of his own price proposal and is determined as a function of the price proposals of other producers. This approach eliminates market power and provides flexibility with discriminatory prices to implement the socially optimal allocation. (2) To ensure uniform efficient price at equilibrium, an individual penalty term is also added to the payments. These penalty terms punish deviations from price proposal of other GenCos. (3) A penalty is imposed if the aggregate production is different from the demand; this penalty ensures market clearing and budget balance. The form of the penalty terms depends on the elasticity of the demand.

2.3.1 Model: Electricity Spot Markets

Based on the model of Section (3.2), we consider a single spot market without constraints. Since we study only a one-time market, for the ease of presentation, we drop the time index and ignore the uncertainty factor in the model of this Section (Also in Sections 2.4 and 2.5). In Section (2.6), we extend the results to the case where both long-term interaction of agents and uncertainty exist.

The social-welfare maximizing allocation for elastic demand is given by the solution of the following problem.

$$\max_{e_n, n\in N} U(\sum_{n\in N} e_n) - \sum_{n\in N} C_n^e(e_n) \qquad \text{(Socially Optimal Dispatch - Elastic)}$$

If the demand is inelastic and equal to $\underline{D}$ then, the social welfare maximizing problem will be

$$\max_{e_n, n\in N} - \sum_{n\in N} C_n^e(e_n) \qquad \text{(Socially Optimal Dispatch - Inelastic)}$$

$$s.t. \quad \sum_{n\in N} e_n \geq \underline{D} \tag{2.25}$$

Note that both of the above problems are strictly convex optimizations with a convex non-empty domain and therefore, have a unique solution.

2.3.2 Design/Mechanism

We present two auction mechanisms, one for elastic and one for inelastic demand. The two mechanisms are only different in their payment methods.

A game form/mechanism is described by $(\mathscr{M}, h)$, where $\mathscr{M}$ is the message/strategy space and $g : \mathscr{M} \to \mathscr{A}$ is the outcome function, a function from the message space $\mathscr{M}$ to the space $\mathscr{A}$ of allocations. Consider the following mechanism.

Message space We choose $\mathscr{M}$ to be

$$\mathscr{M} := (\mathscr{M}_1 \otimes \mathscr{M}_2 \otimes \ldots \otimes \mathscr{M}_N), \tag{2.26}$$

where $\mathscr{M}_n$ is producer n's message space,

$$\mathscr{M}_n := \mathbb{R} \times \mathbb{R}_+, n \in N, \tag{2.27}$$

and $m_n \in \mathscr{M}_n$ is of the form

$$m_n = (\hat{e}_n, \hat{p}_n), \tag{2.28}$$

where $\hat{e}_n$ denotes the amount of good GenCo n proposes to produce, and $\hat{p}_n$ denotes the price GenCo

n proposes to charge per unit of electricity. $\hat{p}_n$ is restricted by $\hat{p}_n \geq 0$.

Allocation Space We choose $\mathscr{A}$ to be

$$\mathscr{A} := (\mathscr{A}_1 \otimes \mathscr{A}_2 \otimes ... \otimes \mathscr{A}_N), \tag{2.29}$$

where $\mathscr{A}_n$ is producer n's allocation space

$$\mathscr{A}_n := [0, x_n] \times \mathbb{R}, n \in N, \tag{2.30}$$

and $a_n \in \mathscr{A}_n$ is of the form

$$a_n = (e_n, r_n), \tag{2.31}$$

where e_n denotes the amount of good GenCo n is scheduled to produce, and r_n denotes the payment to GenCo n.

Outcome function We define $g : \mathscr{M} \to \mathscr{A}$ as follows:

For each $m := (m_1, m_2, ..., m_N) \in \mathscr{M}$ we set

$$g(m) = (e, r) = (e_1, ..., e_N, r_1, ..., r_N), \tag{2.32}$$

where

$$e_n \quad = \quad \hat{e}_n \tag{2.33}$$

For elastic demand the payments are[14]

$$r_n^{elas.} \quad = \quad \hat{p}_{n+1}\hat{e}_n - \hat{p}_n^{-0.5} {\zeta_n^{elas.}}^2 \tag{2.34}$$

$$\zeta_n^{elas.} \quad = \quad D(\hat{p}_{n+1}) - \sum_{n \in N} \hat{e}_n \tag{2.35}$$

$$D(\hat{p}) \quad = \quad (U^{'})^{-1}(\hat{p}) \tag{2.36}$$

$$\hat{p}_{N+1} \quad := \quad \hat{p}_1. \tag{2.37}$$

For inelastic demand the payments are

$$r_n^{inelas.} \quad = \quad \hat{p}_{n+1}\hat{e}_n - (\hat{p}_n - \hat{p}_{n+1})^2 - 2\hat{p}_n {\zeta^{inelas.}}^2 \tag{2.38}$$

$$\zeta^{inelas.} \quad = \quad (\underline{D} - \sum_{n \in N} \hat{e}_n)^+ \tag{2.39}$$

$$p_{N+1} \quad := \quad p_1. \tag{2.40}$$

[14]To make sure the tax is in terms of dollars, appropriate unit conversion factors should be considered in the tax function. To avoid complexities, we have assumed these factors are 1 and have not put them explicitly in the tax terms.

We refer to these auctions for elastic and inelastic demand by $Au^{elas.}$ and $Au^{inelas.}$ correspondingly.

2.3.3 Interpretation of the Design

The key challenges in the design of an efficient mechanism for the problem considered in this Section are (1) dealing with the strategic GenCos' market power; and (2) incentivizing the strategic GenCos to collectively meet the price-taking demand.

Since the mechanism designer, i.e. the ISO, can not alter the producers' cost functions, $C_n^e(.)$, $n = 1, 2, ..., N$, even if he knew their functional form, the only way he could achieve his objective (social welfare maximization) is through the use of appropriate tax incentives/tax functions. For the case of elastic demand, the tax function is $r_n^{elas.} = r_{n,1} + r_{n,2}^{elas.} = [\hat{p}_{n+1}\hat{e}_n] + [-\hat{p}_n^{-0.5}\zeta^{elas.}{}_n^2]$. For the case of inelastic demand, $r_n^{inelas.} = r_{n,1} + r_{n,2}^{inelas.} = [\hat{p}_{n+1,t}e_n] + [-(\hat{p}_n - \hat{p}_{n+1})^2 - 2\hat{p}_n\zeta^{inelas.}{}_t^2]$.

The producers' market power is addressed by offering each producer a price per unit of its production that does not depend on its own price proposal, that is, it is not under its own control. Such an offer induces price-taking behavior among the producers; moreover, allowing discriminatory individual prices charged to GenCos (off-equilibrium) provides the flexibility required for implementing social welfare maximizing generation. One way to achieve these goals is to arbitrarily index the GenCos between 1 and N, and then pay GenCo n according to the price proposal p_{n+1} of GenCo $n+1$, with the convention that $p_{N+1} = p_1$. The term $r_{n,1}$ specifies the amount GenCo n receives for its production e_n in this way.

Incentivizing the producers to bid the same prices, can be achieved by a penalty term for each producer n that depends on his own price. In the same way, incentivizing the strategic producers to collectively meet the demand can be achieved by penalizing over-production or under-production. The term $r_{n,2}^{elas.}$ and $r_{n,2}^{inelas.}$ are such penalty terms that provide both of these incentives and can be thought of as the tax payments the ISO collects from the producers in order to align their productions with the production profile that maximizes the sum of the utilities of producers and the demand. For the inelastic demand, the incentive provided to all producers to bid the same price per unit of capacity is described by the term $-(\hat{p}_{n,t} - \hat{p}_{n+1,t})^2$. The incentive provided to all producers to collectively propose a total production that is equal to the demand, $\underline{D}$, is captured by the term $-2\hat{p}_{n,t}\zeta^{inelas.}{}_t^2$ which is a penalty for underproduction. Moreover, for the elastic demand $r_{n,2}^{elas.}$[15] ensures the equilibrium price is equal to the marginal utility of the demand at the total amount of production. Under this price all the electricity supplied at the suggested price is bought and used by the demand. Note that $D(p)$ is the optimal demand for price p, because it is the solution of the

[15]Following the above guidelines, different penalty terms can be designed for efficient auctions and those terms proposed here are just one possible formulation. Note that the terms $r_n^{elas.}$ and $r_n^{inelas.}$ should be concave to ensure the results hold.

optimization problem below:

$$D(p) \in \arg\max_{d \geq 0} U(d) - pd. \tag{2.41}$$

Note that, since the above optimization is strictly convex, it has a unique solution.

2.3.4 Comparison of the Elastic and Inelastic Static Auctions

We now contrast the two auction mechanisms we proposed for elastic and inelastic divisible good. The key differences appear in the specification of the tax function which provides incentives to strategic GenCos to align their interests/goals with those of the ISO. In the case of elastic demand, the money the demand pays defines the electricity amount it consumes. Thus, this price should be less than or equal to marginal utility of consumption at the amount injected to the market. This expresses the budget balanced property the mechanism should have. In the case of inelastic demand, budget balanced is not required because the demand pays to the GenCos any amount it is charged. In return, the GenCos should satisfy an inequality constraint of the form $\sum_{n=1,\ldots,N} e_n \geq \underline{D}$. The difficulty in this case is designing a price efficient mechanism (a mechanism where the price is limited by marginal cost of production) which also implements $\sum_{n=1,\ldots,N} e_n \geq \underline{D}$.

As a result of these differences, the two mechanisms are different in the way GenCos contribute to the tax payment that is imposed so as to achieve a feasible allocation at equilibrium.

2.3.5 Properties

The properties possessed by the two proposed mechanisms are described in this Section. We consider Nash equilibrium (NE) as the solution concept. We denote the equilibrium messages and outcomes corresponding to GenCo $n \in N$ by $\hat{p}_n^*, \hat{e}_n^*, r_n^{elas.*}, r_n^{inelas.*}, p_n^*$ and e_n^*.

We first define trivial equilibrium (Definition 2.3.1); we show that the auction for inelastic demand has exactly one trivial NE (Lemma 2). Hereafter, for simplicity of presentation whenever we refer to NE, we mean non-trivial NE unless otherwise stated. We prove that at any equilibrium (trivial or non-trivial) outcome of the game induced by the above mechanisms, supply and demand meet each other (Lemma 1). We use Lemmas 2 and 1 to prove price efficiency, budget balanced, social welfare maximization, and individual rationality at any Nash equilibrium in Theorems 1-4 respectively. The proof of all these properties is presented in Appendix A. We conclude this Section by discussing uniqueness of the NE, Pareto dominance of the non-trivial NE compared to the trivial NE for the inelastic demand, and the anonymity of the mechanisms.

Definition 2.3.1 (TRIVIAL EQUILIBRIUM)**.** *An equilibrium where price bids $\hat{p}_n$ are zero for all $n \in N$ is defined to be a trivial equilibrium.*

Lemma 1 (MARKET CLEARANCE)**.** *At every non-trivial equilibrium, supply meets demand, i.e. for elastic demand*

$$\zeta_n^{elas.*} = 0, \tag{2.42}$$

$$D(\hat{p}_{n+1}^*) - \sum_{n \in N} \hat{e}_n^* = 0, \tag{2.43}$$

for all $n \in N$ and for inelastic demand

$$\zeta^{inelas.*} = 0, \tag{2.44}$$

$$(\underline{D} - \sum_{n \in N} e_n^*) \geq 0. \tag{2.45}$$

Theorem 1 (PRICE EFFICIENCY)**.** *The auctions proposed for elastic and inelastic demand are price efficient at any non-trivial equilibrium; that is there is a uniform price equal to the marginal cost of the next one unit of production by producers, i.e. for every producer $n \in N$,*

$$\hat{p}_n^* = \hat{p}_{n+1}^* = \hat{p}^*, \tag{2.46}$$

$$r_n^* = \hat{p}^* \hat{e}_n^*, \tag{2.47}$$

$$\left. \frac{\partial r_n}{\partial \hat{e}_n} \right|_{m^*} = \hat{p}^*, \tag{2.48}$$

$$\hat{p}^* = C'(\hat{e}_n^*). \tag{2.49}$$

Theorem 1 shows that at equilibrium all producers propose the same price per unit of electricity (Eq. 2.46). This equilibrium price is both the average price (Eq. 2.47) and the marginal price (Eq. 2.48), and is efficient in the sense that it is equal to the marginal cost of production for non-saturated producers (Eq. 2.49). This Theorem shows that the proposed mechanisms incentivize producers to reveal at equilibrium their true cost of producing the next unit of electricity provided that they still have free capacity. This marginal cost is the same across all producers with free capacity, and therefore, would be the marginal cost of production of a hypothetical entity who would own all the producers.

Theorem 2 (BUDGET BALANCED)**.** *The proposed auctions are budget balanced at any trivial or non-trivial equilibrium. For elastic demand, if $\sum_{n \in N} \hat{e}_n^* > 0$, this means the equilibrium price is equal to marginal utility of the demand,*

$$\hat{p}^* = U'(\sum_{n \in N} \hat{e}_n^*). \tag{2.50}$$

Note that in contrast to the auction for elastic demand, budget balance is always satisfied in the auction for inelastic demand by charging demand the total payment to the GenCos.

Theorem 3 (SOCIAL OPTIMALITY)**.** *At any non-trivial equilibrium of the proposed auction for elastic or inelastic demand, the allocation of the demand to GenCos is socially optimal according to the unique solution of Problem (Socially Optimal Dispatch - Elastic) for allocating an elastic demand and Problem (Socially Optimal Dispatch - Inelastic) for allocating an inelastic demand.*

Theorem 1 determines the price bids and Theorem 3 determines the production bids at equilibrium. Using these bids, we can establish individual rationality.

Theorem 4 (INDIVIDUAL RATIONALITY)**.** *The proposed auction mechanisms for elastic and inelastic demand are individually rational, that is at any trivial or non-trivial NE of the game induced by the mechanisms and for every GenCo $n \in N$, the corresponding allocation $(e_n^*, r_n^{elas.*})$ or $(e_n^*, r_n^{inelas.*})$ is weakly preferred to the allocation $(0,0)$ (the allocation a producer receives when it does not participate in the market).*

Theorem 4 shows that each strategic producer voluntarily participates in the electricity pooling market.

Lemma 2 (UNIQUENESS OF TRIVIAL NE)**.** *The game induced by the proposed auction for inelastic demand has exactly one trivial Nash equilibrium, where for all $n \in N$, $p_n^* = 0$ and $e_n^* = 0$. The game induced by the auction for elastic demand does not have any trivial equilibrium if the solution to Problem (Socially Optimal Dispatch - Elastic), e_n^* is not zero for all $n \in N$, otherwise it has a unique trivial equilibrium where for all $n \in N$, $p_n^* = 0$ and $e_n^* = 0$.*

Remark 2.3.1. *[UNIQUENESS] The games induced by the proposed mechanisms have a unique non-trivial NE. This is because by Theorem (3), the bids for electricity production at equilibrium are corresponding to the unique solution of Problem (Socially Optimal Dispatch - Elastic) for elastic demand and Problem (Socially Optimal Dispatch - Inelastic) for inelastic demand. Furthermore, by Theorem (1), the price bids of all the GenCos is the same at equilibrium, and equal to marginal cost of producing next unit of electricity.*

[PARETO DOMINANCE] Since $(0,0)$ is the outcome of the trivial NE of the game induced by the mechanism, Theorem 4 shows that the non-trivial NE Pareto dominates the trivial NE.

[ANONYMITY] The proposed mechanism is anonymous at equilibrium. In other words, the equilibrium outcome does not depend on the index order of the GenCos[16].

[16] To see this note that the payments and allocations at equilibrium do not depend on the producers' indices rather as shown by Theorems (3), at equilibrium the electricity produced by each GenCo is according to the social welfare maximizing production which is independent of the indexing. Furthermore, as shown in Lemma (1) (Eq. (2.46)) the price paid per unit of production at equilibrium is the same for all the producers and independent of the producers' indexing, and the terms $r_{n,2}^{elas.}$ and $r_{n,2}^{inelas.}$, $\quad n = 1,2,...,N$, in the tax function are equal to zero (Eqs. 2.47).

Anonymity at off-equilibrium outcomes is also a desired feature for markets. For the case of elastic demand, we present another mechanism that is anonymous both at equilibrium and off equilibrium by using the following tax payments that are independent of the GenCos' indexing.

$$r_n^{elas.} = \hat{p}_{-n}\hat{e}_n - \hat{p}_n^{-0.5}\zeta_n^{elas.2} \tag{2.51}$$

$$\zeta_n^{elas.} = D(\hat{p}_{-n}) - \sum_{n\in N}\hat{e}_n \tag{2.52}$$

$$\hat{p}_{-n} = \frac{\sum_{j\in N, j\neq n}\hat{p}_j}{N-1}. \tag{2.53}$$

The mechanism described by Eqs. (2.26)-(2.28), (2.33), and (2.51)-(2.53) is an efficient auction that is also anonymous both at equilibrium and off equilibrium.

2.4 Efficient Auction with Individual Constraints

We consider auctions where sellers have convex individual constraints. We show that with convex individual constraints, the auctions proposed in Section (2.3.2) remain efficient without any need for additional prices (and corresponding markets for discovering prices) or other instruments.

We will study these auctions within the context of GenCo's capacity expansion without reliability constraints. In this Section and Sections 2.5 and 2.6, we focus on the case of elastic demand; the extension of techniques and results to corresponding problems with inelastic demand can be achieved in a similar way. We design an efficient capacity-trade auction, and show that the proposed auction is efficient (Theorem 5). Moreover, the GenCos use all of the expanded capacity in the spot market and the spot market price covers their investment cost (Theorem 6).

At the end of the section, we interpret the results within the context of supply chains. We will also interpret the results specifically for energy-only electricity markets in Section (2.7).

2.4.1 Model: Investment without Reliability Constraint

Consider the model of Section (3.2) restricted to the following capacity-trade competition: GenCos individually decide on their investment at time 1 and then the demand is auctioned at the spot market of time 1^+. GenCos' production is limited by their individual capacity and their capacity expansion is limited by their individual expansion limits. The social-welfare maximizing allocation for elastic demand is a solution to the problem

$$\max_{e_n,\Delta x_n, n\in N} U(\sum_{n\in N} e_n) - \sum_{n\in N} C_n^e(e_n) - \sum_{n\in N} C_n^x(\Delta x_n) \qquad \text{(Socially Optimal - Individual Const.)}$$

$$s.t. \quad 0 \le e_n \le x_{n,0} + \Delta x_n \quad \forall n \in N \tag{2.54}$$

$$0 \le \Delta x_n \le \overline{\Delta x_n} \quad \forall n \in N \tag{2.55}$$

Problem (Socially Optimal - Individual Const.) is a strictly convex optimization problem with a convex non-empty domain; therefore it has a unique solution.

2.4.2 Design/Mechanism

We propose the following dynamic mechanism for capacity-trade competition. GenCos simultaneously decide on their expansion Δx_n at time 1. Then at time 1^+, they participate in the efficient auction (spot market) for elastic demand proposed in Section (2.3.2). Note that due to constraint in Eq. (2.54), GenCo n's message space in the spot market is limited by $0 \le \hat{e}_n \le x_n$.

2.4.3 Properties

We consider subgame perfect Nash equilibrium (SPNE) as the solution concept and study the equilibrium properties of our proposed capacity-trade mechanism. We denote the equilibrium outcome of the game induced by the mechanism by $\Delta x_n^*, e_n^*$ and p_n^*. The properties of the mechanism are summarized by Theorems 5 and 6. The proof of all the Theorems in this Section is presented in Appendix B.

Theorem 5 (EFFICIENCY)**.** *The game induced by the proposed mechanism for capacity-trade competition (capacity expansion and production) has unique sub-game perfect Nash equilibrium. This unique equilibrium (i) implements socially efficient investment and production corresponding to Problem (Socially Optimal - Individual Const.), (ii) is individually rational, (iii) results in efficient price in the spot market and (iv) is budget balanced.*

GenCos cover their cost of investment when their capacity becomes saturated and therefore, the market price is above their marginal cost of production. In the following Theorem, we show that the GenCos' investment on capacity is always covered, and it is covered at its marginal cost.

Theorem 6 (SATURATION & SCARCITY PRICE)**.** *If GenCo n's expansion at time* 1 *is positive, i.e.* $\Delta x_n^* > 0$, *then its production in the spot market at time* 1^+ *is saturated i.e.*

$$e_n^* = x_{n,0} + \Delta x_n^*. \tag{2.56}$$

Moreover, the equilibrium price in the spot market covers the investment at its marginal cost, i.e.

$$p^* = C_n^{e\prime}(e_n^*) + C_n^{x\prime}(\Delta x_n^*) \text{ } \mathit{if} \text{ } 0 < \Delta x_n < \overline{\Delta x}_n. \tag{2.57}$$

2.4.4 Interpretation

Our results for auctions with convex individual constraints can be interpreted in a more general framework within the context of supply chains. Consider a supply chain for an end-product. The producers are competing in the end-product market to meet the demand, but there is no competition on the other upstream goods/resources in the supply chain; each producer has only individual costs and limits in the purchase/preparation of the upstream goods/resources which are independent of the other producers' purchase/preparation of those goods/resources. Our results on auctions with individual constraints show that implementation of the socially optimal outcome in this setup can be achieved by a single efficient market for the end-product and there is no need to design markets for the upstream goods/resources. In other words, our efficient market for the end-product not only achieves socially optimal allocations of the end-product demand to the producers, but also sends back correct signals for efficient investment on all the other upstream goods/resources in the supply chain.

2.5 Efficient Auction with Homogeneous (Additive) Joint Constraints

We consider auctions with homogeneous (additive) joint constraints. Joint constraints are public (uneconomic) constraints and their implementation requires appropriate tax/subsidy. We introduce a technique for decomposing homogeneous joint constraints into a sequence of individual constraints and then implementing them by a sequence of efficient auctions with individual constraints. The equilibrium price in the market corresponding to each homogeneous joint constraint determines the efficient tax/subsidy needed to implement that constraint.

We present our decomposition technique and results on homogeneous joint constraints through the study of reliability constraints (planning reserve requirements) for electricity investment. We also discuss in Section (2.7) how to implement carbon emission constraints and lines' thermal capacity constraints using our results on auctions with homogeneous joint constraints.

2.5.1 Model: Electricity Investment with Reliability Constraint

We use the model of Section (3.2) and add the reliability constraint to the model of Section (2.4.1). The social welfare maximizing investment and production are a solution to the Problem (Socially Optimal - Joint Const.)

$$\max_{e_n,\Delta x_n,n\in N} U(\sum_{n\in N} e_n) - \sum_{n\in N} C_n^e(e_n) - \sum_{n\in N} C_n^x(\Delta x_n) \qquad \text{(Socially Optimal - Joint Const.)}$$

$$s.t. \quad 0 \le e_n \le x_{n,0} + \Delta x_n \quad \forall n \in N \tag{2.58}$$

$$0 \le \Delta x_n \le \overline{\Delta x}_n \quad \forall n \in N \tag{2.59}$$

$$\sum_{n\in N} (x_{n,0} + \Delta x_n) \ge \underline{x} \tag{2.60}$$

Problem (Socially Optimal - Joint Const.) is a strictly convex optimization with a convex non-empty domain; therefore, it has a unique solution.

2.5.2 Design/Mechanism

We decompose Problem (Socially Optimal - Joint Const.) and implement its solution with separate markets. The decomposition and implementation method presented here is for reliability constraints but the method can be used for problems with other convex homogeneous joint constraints.

Decomposition: Consider $x = (x_1, x_2, ..., x_N)$ where $x_n = x_{n,0} + \Delta x_n$. Problem (Socially Optimal - Joint Const.) is equivalent to the following two sequential optimization

$$\max_{\Delta x_n,n\in N} V^+(x) - \sum_{n\in N} C_n^x(\Delta x_n) \qquad \text{(Joint Const. Decompos. 1)}$$

$$s.t. \quad 0 \le \Delta x_n \le \overline{\Delta x}_n \quad \forall n \in N \tag{2.61}$$

$$\sum_{n\in N} x_n \ge \underline{x}, \tag{2.62}$$

followed by

$$V^+(x) = \max_{e_n,n\in N} U(\sum_{n\in N} e_n) - \sum_{n\in N} C_n^e(e_n) \qquad \text{(Joint Const. Decompos. 2)}$$

$$s.t. \quad 0 \le e_n \le x_n \forall n \in N. \tag{2.63}$$

Market Design: Inspired by the above decomposition, we propose the following dynamic market design. First at time 1, the amount of capacity required for reliability, $\underline{x}$, is auctioned among the GenCos using efficient inelastic *capacity market*. The design of this market is the same as the efficient inelastic spot market proposed in Section (2.3). Next, at time 1^+, the elastic demand $U(d)$

is auctioned among the GenCos using an efficient elastic spot market that is the same as that of Section (2.3). Note that due to constraint in Eq. (2.63), the GenCo n's message space in the spot market is limited by $0 \leq \hat{e}_n \leq x_n$, and in the same way, due to constrain in Eq. (2.61), its expansion proposals in the capacity market should be between zero and $\overline{\Delta x}_n$.

We prove the properties of this dynamic market in the next section.

2.5.3 Properties

The properties of the proposed dynamic auction for implementing reliability constraints are summarized by Theorems (7) and (8). In Section (2.6.3), we extend these properties to the more general case where GenCos have long-term interaction with symmetric uncertainty and prove them in Appendix C.

Theorem 7 (EFFICIENCY)**.** *The mechanism proposed for auctions with homogeneous joint constraints is individually rational, budget balanced, and price efficient both in the capacity market and in the spot market. Furthermore, if the solutions of Problem (Socially Optimal - Joint Const.) results in non-negative total expansion i.e.* $\sum_{n1N} \Delta x_n^* 0$*, then the game induced by the proposed mechanism has a unique non-trivial SPNE with the allocations corresponding to the optimal solutions of Problem (Socially Optimal - Joint Const.). The game induced by the proposed mechanism has also a trivial SPNE where the investment on capacity expansion* Δx_n*,* $n \in N$*, are equal to zero. The non-trivial SPNE Pareto dominates the trivial SPNE.*

We present a sketch of the proof for Theorem (7) in three steps (the complete proof is provided in Appendix C): First, at time 1^+ for all vales of $x = \{x_1, x_2, ..., x_n\}$, the efficient elastic spot market implements the solution of Problem (Joint Const. Decompos. 2) and is budget balanced, individually rational and price efficient (Theorems (5)). Next, we prove the equilibrium in the capacity market is the solution of Problem (Joint Const. Decompos. 1) on two steps. Define $\tilde{V}_n^+$ to be the GenCo n's utility in the spot market:

$$V_n^+(x) = -C_n^e e_n^*) + r_n^*. \tag{2.64}$$

In the inelastic capacity market of time 1, each GenCo wants to maximize his individual utility which is $V_n^+(x) - C_n^x(\Delta x_n)$ as a best response to the other GenCos' expansion decisions, $(\Delta x_1, \Delta x_2, ...\Delta x_{n-1}, \Delta x_{n+1}, ..., \Delta x_n)$. Therefore, by Theorem (5), the equilibrium of the efficient capacity market is budget balanced, individually rational and price efficient, and implements the solution of the following optimization problem:

$$\max_{x_n, n\in N} \sum_{n\in N} [V_n^+(x) - C_n^x(\Delta x_n)] \qquad \text{(Joint Const. Decompos. 3)}$$

$$s.t. \quad 0 \le \Delta x_n \le \overline{\Delta x}_n \quad \forall n \in N \tag{2.65}$$

$$\sum_{n\in N} x_n \ge \underline{x}. \tag{2.66}$$

To complete the proof, we show that the solution of Problem (Joint Const. Decompos. 3) is equivalent to the solution of Problem (Joint Const. Decompos. 1). This is because these solutions satisfy the same first order optimality constraints considering that for each $n \in N$, V^+ and V_n^+ have the same derivatives with respect to each Δx_n. Steps one to three show that the the proposed dynamic markets implement Problem (Socially Optimal - Joint Const.) in NE and are budget balanced, individually rational, and price efficient.

Next, we can comparing the mechanism for homogeneous joint constraint with the mechanism for auction with individual constrains proposed in Section (2.4). In contrast to auctions with individual constraints, saturation and scarcity price results of Theorem 6 do not necessarily hold for our auctions with homogeneous joint constraints. In other words, it is possible that a GenCo expands its capacity without using the expanded capacity in the spot market. However, the price in the capacity market is efficient (Theorem (8)).

Theorem 8 (EFFICIENT SUBSIDY/TAX)**.** *At the non-trivial SPNE, the equilibrium price in the capacity market, p^{c*}, is equal to the efficient level of subsidy/tax per unit of capacity expansion needed to implement the joint constraint in Eq. (2.60), that is*

$$p^{c*} = C_n^{x\prime}(\Delta x_n) + C_n^{e\prime}(e_n) - p^* \tag{2.67}$$

where p^* is the equilibrium price in the spot market. This equilibrium price in the capacity market goes to zero if the joint constraint is not binding.

Remark 2.5.1 (MULTIPLE JOINT CONSTRAINTS)**.** *Our design by decomposition technique can be extended to multiple convex homogeneous joint constraints by considering one price signal for each constraint which is the efficient tax/subsidy for the corresponding joint constraint. Each efficient price can be discovered using a separate efficient market.*

Remark 2.5.2 (ELASTIC JOINT CONSTRAINT)**.** *Instead of the hard constraint for reliability in Eq. (2.66), some proposals assume the market operator has an elastic value function for reliability which is an increasing function of the existing capacity, $U^r(\sum_{n\in N} x_n)$. Our decomposition technique can be extended to work for this case. The corresponding capacity market will be an efficient auction of an elastic divisible good with utility $U^r(.)$.*

2.6 Long-Term Interaction with Common Uncertainty

In this Section, we extend the previous results in two directions: considering long-term interaction of agents and adding common uncertainty. Long-term interaction may result in collusion among competing agents; furthermore, agents should take into account the effect of uncertainty in their utility. We consider risk-neutral agents who should maximize their expected utility. We show that by using a sequence of the auction mechanisms presented in previous sections, we can implement the efficient outcome with the additional desirable features.

We present the results of this section within the context of long-term electricity investment with reliability constraints. In practice, GenCos interact for multiple years and have uncertainty about future demand, cost of production, cost of expansion, etc. We model this environment and design an efficient dynamic auction mechanism for it by repeating the mechanism proposed in Section (2.5.2) in each year.

In order to study the properties of this mechanism, we first show that in all of the sub-game perfect Nash equilibria (SPNE) of the game induced by it, GenCos' strategies are only a function of the payoff-relevant information i.e. existing capacities and realization of past uncertainties (Lemma 3); an immediate consequence is that GenCos act myopically in the electricity spot markets, i.e. maximizing their immediate outcome (Corollary 2.6.1). We use Lemmas (3) to prove that all SPNEs of the game induced by the mechanism are individually rational, budget balanced and price efficient; moreover, the unique non-trivial SPNE Pareto dominates all trivial SPNEs and is socially optimal (Theorem 9). For infinite horizon interaction, there may exist non-Markov equilibria i.e. Lemma 3 does not hold, but by considering the sub-set of equilibria where the strategies are Markov and a only a function of pay-off relevant information, we can prove the extend the same results to infinite horizon interaction (Theorem 10).

2.6.1 Model: Long Term Investment with Reliability Constraints

We extend the model of Section (2.5) for investment with reliability constraint to the case where GenCos interact over a finite time horizon and have common uncertainty that is common knowledge among them. The expected social welfare maximizing investment and production is

$$\max_{\sigma_{n,t}\in\Sigma_{n,t},\gamma_{n,t}\in\Gamma_{n,t}:n\in N,t\in T} E_{\mathbf{w}_{1:T}}\Bigg[\sum_{t\in T}\beta^t U_t(\sum_{n\in N} e_{n,t},\mathbf{w}_t)-$$

(Socially Optimal - Longterm Uncertainty)

$$\sum_{t\in T}\sum_{n\in N}\beta^t\Big\{C^e_{n,t}(e_{n,t},\mathbf{w}_t)+C^x_{n,t}(\Delta x_{n,t},\mathbf{w}_{t-1})\Big\}+\beta^{T+1}\sum_{n\in N}\eta(\mathbf{w}_T)x_{n,T}\Bigg] \tag{2.68}$$

$$s.t.\quad 0\le e_{n,t}\le x_{n,0}+\sum_{\tau=1,\ldots,t}\Delta x_{n,\tau} \tag{2.69}$$

$$0\le \Delta x_{n,t}\le \overline{\Delta x}_{n,t}(\mathbf{w}_{t-1}) \tag{2.70}$$

$$\sum_{n\in N} x_t \ge \underline{x_t}(\mathbf{w}_{t-1}) \tag{2.71}$$

In Lemma (3), we prove that the solution of Problem (Socially Optimal - Longterm Uncertainty), the expansion strategies at time t, $\sigma^*_{n,t}$, are only functions of $S_t=\{x_{t-1},w_{1:t-1}\}$ and generation strategies at time t^+, $\gamma^*_{n,t}$, are only functions of $S_{t^+}=\{x_t,w_{1:t}\}$. In other words, S_t and S_{t^+} are information state for Problem (Socially Optimal - Longterm Uncertainty). We will also prove that the same is true for the GenCo's SPNE strategy in the game induced by the mechanism we design.

2.6.2 Design/Mechanism

Using dynamic programming and value function, Problem (Socially Optimal - Longterm Uncertainty) can be decomposed into a set of sequential optimization problems, one for each year (See Appendix (C.1)). Furthermore, using the decomposition technique of Section (2.5.2), the optimization of each year $t\in T$ can be further decomposed into two optimization problems, one for capacity expansion and one for electricity generation. Considering this decomposition, we use $2\times T$ markets for implementing the corresponding outcome: each year $t\in T$ has a capacity market for auctioning the inelastic capacity $(\underline{x}_t(\mathbf{w}_{t-1})-\sum_{n\in N}x_{n,t-1})$[17], followed by an elastic spot market for auctioning electricity demand with utility $U_t(d_t,w_t)$. The timeline of events for year t is shown in Fig. (2.4).

2.6.3 Properties

In the game induced by the energy-and-capacity mechanism, the strategy of each seller n consists of both strategies $\tilde{\sigma}_{n,t}, t\in T$ determining their bids in the capacity market and strategies $\tilde{\gamma}_{n,t^+}, t\in T$

[17] We assume the level of expansion required for reliability is always non-negative. Example of cases where it can become negative is when the demand prediction suddenly falls (due to technological changes) and the level of capacity required for reliability can fall below existing capacity. Moreover, we assume the reliability constraints are always binding; otherwise the GenCos would like to invest more than the capacity allocated to them in the capacity market. Changing these assumptions does not change the results, but they make presentation of the results complicated.

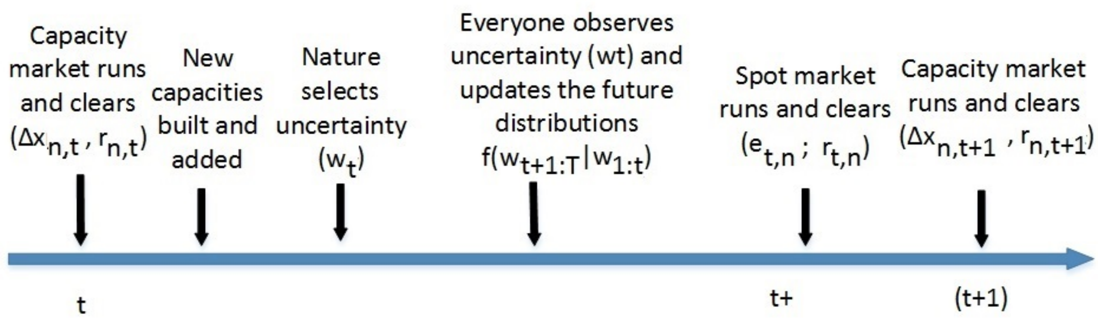


Figure 2.4: Timeline of events in year t for the proposed markets of Section (2.6.2).

determining their bids in the spot markets. The history of the game at the time of expansion is

$$\tilde{h}_t = \{x_0, \Delta\tilde{x}_{1:t-1}, \tilde{p}_{1:t-1}, \hat{e}_{1:t-1}, \hat{p}_{1:t-1}, w_{1:t-1}\} \in \tilde{H}_t \tag{2.72}$$

$$\tilde{H}_t := \{x_0, \Delta\tilde{x}_{1:t-1}, \tilde{P}_{1:t}, \hat{E}_{1:t-1}, \hat{P}_{1:t-1}, w_{1:t-1}\} \tag{2.73}$$

$$\tilde{\sigma}_{n,t} : \tilde{H}_t \to [0, \overline{\Delta x}_{n,t}] \times \mathbb{R}_+ \tag{2.74}$$

And at the spot market t it is

$$\tilde{h}_{t^+} = \{x_0, \Delta\tilde{x}_{1:t}, \tilde{p}_{1:t}, \hat{e}_{1:t-1}, \hat{p}_{1:t-1}, w_{1:t}\} \tag{2.75}$$

$$\tilde{H}_{t^+} = \{x_0, \Delta\hat{x}_{1:t}, \tilde{P}_{1:t}, \hat{E}_{1:t-1}, \hat{P}_{1:t-1}, w_{1:t}\} \tag{2.76}$$

$$\tilde{\gamma}_{n,t^+} : \tilde{H}_{t^+} \to [0, x_{n,t}] \times \mathbb{R}_+ \tag{2.77}$$

We denote the set of all such strategies by $\tilde{\Sigma}_{n,t}$ and $\tilde{\Gamma}_{n,t}$.

The properties of the proposed energy-only economy at its SPNE are described by Lemma 3 and Theorems 9-10. The proof of all these properties is presented in Appendix C.

We first show that strategies at any SPNE of the game induced by the energy-and-capacity mechanism are only a function of the payoff relevant information i.e. they are Markov.

Lemma 3. *(MARKOV STRATEGIES) At any solution of Problem (Socially Optimal - Longterm Uncertainty) or any SPNE of the game induced by the proposed finite-horizon mechanism of Section (2.6.2), expansion strategies of all GenCos $n \in N$ at time t are functions of $S_t = \{x_{t-1}, w_{1:t-1}\}$ and the production strategies at time t^+ are functions of $S_{t^+} = \{x_t, w_{1:t}\}$.*

Lemma (3) implies the following corollary.

Corollary 2.6.1 (MYOPIC BEHAVIOR)**.** *Lemma 3 implies that the GenCos act myopically in the spot market. This is because the actions of the GenCos in the spot market at time t^+, production and price proposals, do not change the information state S_τ and S_{τ^+} for $\tau > t$ i.e. these actions* *do not have any effect in the future strategies of the agents. Therefore, it is best for GenCos to act myopically and maximize their utility in the current spot market.*

We call a SPNE of the game induced by the mechanism non-trivial, if the equilibrium price proposal in the capacity markets of time t, p_t^{c*} is not zero for all $t \in T$. Theorem (9) shows the game induced by the mechanism has a unique non-trivial SPNE which Pareto dominates all the trivial SPNEs.

Theorem 9 (EFFICIENCY)**.** *There is a unique non-trivial SPNE that Pareto dominates all the trivial SPNEs of the game induced by the mechanism of Section (2.6.2) and has the following features.*

(i) [SOCIAL OPTIMALITY] The expansion and production outcomes corresponding to the unique non-trivial SPNE are socially optimal, that is they are equal to the solution of Problem (Socially Optimal - Longterm Uncertainty).

(ii) [PRICE EFFICIENCY] For all $t \in T$, there is a unique capacity price proposal, p_t^{c*} in the capacity market, and a unique electricity price, p_t^* in the spot market for all GenCos $n \in N$ which are efficient, that is

$$p_t^* = C'(e_{n,t}^*) \text{ if } 0 < \hat{e}_{n,t}^* < x_{n,t}, \tag{2.78}$$

$$p_t^{c*} = C^{x\prime}(\Delta x_{n,t}^*) - E_{\mathbf{w}_t|S_t}\left\{ \frac{\partial V_{n,t^+}(x_{t^+}, w_{1:t-1}, \mathbf{w}_t)}{\partial \Delta x_{n,t}} \bigg|_{m_t^*} \right\} \text{ if } 0 < \Delta x_{n,t}^* < \overline{\Delta x}_{n,t}, \tag{2.79}$$

where $V_{n,t^+}(S_{t^+})$ is the future expected revenue of producer n from time t^+ on.

(iii) [BUDGET BALANCED] All the capacity and spot markets are budget balanced at equilibrium. For the elastic spot market at time $t \in T$,

$$\hat{p}_t^* = \frac{\partial U_t(\sum_{n\in N} e_{n,t}, w_t)}{\partial(\sum_{n\in N} e_{n,t})}. \tag{2.80}$$

if $\sum_{n\in N} e_n > 0$.

(iv) [INDIVIDUAL RATIONALITY] The proposed mechanism is individually rational, that is for every time $t \in T$ and realization of the history, the sum of the current and expected future utility is non-negative.

Note that Eq. (2.79) is a generalization of Eq. (2.67) and shows that the price in the capacity market is equal to the difference between the marginal cost of expansion and the expected marginal future stream of revenue. This is exactly the missing money gap that is covered by the capacity market. As this gap goes to zero, the price for capacity also goes to zero.

Other than the unique efficient SPNE, the game induced by the proposed mechanism has a number of trivial SPNEs. Using the intuition in Lemma (2), this is because at each capacity market at time t, and history $\tilde{(h)}_t$, $t \in T$, for auctioning inelastic capacity market, the GenCos may choose one of the following two strategies: trivially bid zero capacity expansion and zero capacity price or propose the socially optimal capacity expansion and the corresponding efficient price.

The above results can not be directly extended to problems where GenCos interact for infinite horizon in general, because the equilibrium strategies may be non-Markov, and therefore, GenCos may collude (Lemma (3) does not hold necessarily). However, the result for finite horizon can be extended to infinite horizon with some extra assumptions.

To extend the model to infinite horizon, assume the random variables $w_t, \quad t \in T$ are Markovian, i.e.

$$f_{w_t}(w_t|w_{1:t-1}) = f_{w_t}(w_t|\mathbf{w}_{t-1}) \tag{2.81}$$

We use this assumption to focus on stationary strategies. In the following theorem, we prove that by focusing on Markov equilibria (where strategies that are only a function of $S_t^{inf} = \{x_{t-1}, \mathbf{w}_{t-1}\}$ and $S_{t^+}^{inf} = \{x_t, \mathbf{w}_t\}$) the results can be extended to the infinite horizon interaction.

Theorem 10 (INFINITE HORIZON INTERACTION)**.** *The game induced by the infinite horizon mechanism has a unique non-trivial Markov perfect equilibrium in which strategies are only a function of the payoff relevant information $S_t^{inf} = \{x_{t-1}, \mathbf{w}_{t-1}\}$ and $S_{t^+}^{inf} = \{x_t, \mathbf{w}_t\}$. This equilibrium is socially optimal, price efficient, individually rational and budget balanced. Moreover it Pareto dominates all trivial Markov perfect equilibria.*

2.7 Implementing Electricity Policy Targets

We use the results the of previous sections to design reliable, sustainable, and price efficient electricity markets considering the links thermal constraints. We propose two sets of markets: energy-and-capacity and energy-only. We also discuss how our results contribute to electricity policy debates mentioned in Section (2.1).

2.7.1 Design/Mechanism

We describe how to design a mechanism that implements the solution of Problem (Socially Optimal) stated in Section (3.2), and in addition, is individually rational, budget-balanced, and price efficient. In Problem (Socially Optimal), we have individual constraints, expressed by Eqs. 2.16), (2.17), (2.21) and (2.22), a homogeneous joint constraints which are the reliability constraint of Eq. (2.18) and the lines' thermal capacity of Eq. (2.20), and non-homogeneous joint constraints that describe the carbon emission limit, Eq. (2.19). Since Eqs. (2.18), (2.19), and (2.20) are all in terms of productions at time t, in order to decompose them into separate optimization problems, we introduce carbon permits and line transmission rights as auxiliary variables. Carbon permit is also required since Eq. 2.19 is non-homogeneous.

Let auxiliary variable $b_{n,t}$ be GenCo n's carbon emission permit at time t; then Eq. (2.19) can be replaced with one individual and one homogeneous joint constraint as follows

$$0 \leq e_{n,t} \leq b_{n,t}/\alpha_{n,t}(\mathbf{w}_{t-1}) \tag{2.82}$$

$$\sum_{n \in N} b_{n,t} \leq \overline{B}_t(\mathbf{w}_{t-1}). \tag{2.83}$$

Also, let auxiliary variables $\bar{I}_{n,uv,t}$ to be GenCo n's transmission right permit at time t in the direction of u to v. Then we can replace Eq. (2.20) with one individual and one homogeneous joint constraint as follows:

$$e_{n,vu,t} \leq \bar{I}_{n,uv,t}, \tag{2.84}$$

$$\sum_{n \in N} \bar{I}_{n,uv,t} \leq \bar{I}_{uv}. \tag{2.85}$$

Then, Problem (Socially Optimal) with constraints and replaced by and , respectively, if of the same form as Problem Socially Optimal - Joint Const. of Section (2.5); therefore, we can use the same mechanism design methodology (decomposition technique) to obtain a mechanism that possesses the properties established by Theorems 7 and 8.

2.7.1.1 Energy-and-Capacity Markets

Using the modifications of Eqs. (2.19) and (2.20) by the constraints of (2.82)-(2.83) and (2.84)-(2.85), one way to implement electricity policies is to use one market for each homogeneous joint constraint in the modified problem. The set of markets at each time t will be the following: one elastic spot market, one inelastic capacity market, one inelastic carbon permit market, and $2 \times V$ inelastic transmission right markets (one market for each link in each direction). All of these markets are auctions of divisible goods and can be designed by the method proposed in Section (2.3). Carbon permit and transmission right markets are used to implement a cap constraint, while the capacity market is for implementing a floor constraint. Therefore, carbon permit and transmission right markets collect money from GenCos resulting in higher prices for electricity in the spot market compared to when the carbon permit and transmission right markets do not exist; while the capacity market pays to the sellers resulting in lower prices in the spot market compared to when the capacity market does not exist. The time-line for this dynamic capacity-and-energy mechanism is shown in Figure 2.5. Note that in this design, there are two transmission right markets for each line, one

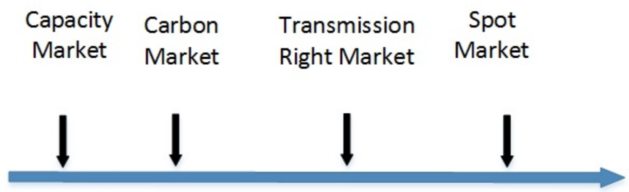


Figure 2.5: Timeline for the proposed capacity-and-energy mechanism

for each direction, hence two prices. GenCos can bid any positive or negative transmission rights in transmission right markets; since the transmission right is the maximum amount of electricity that can be sent in a specific direction, a negative transmission right means the GenCo must send a power flow in the opposite direction and the amount of this flow must be larger than the absolute value of the right. If some GenCos' are allocated negative transmission rights, then other GenCos can be allocated positive transmission right even greater than $\bar{I}_{uv}$ as far as Eq. (2.85) is satisfied. The above design implements Problem (Socially Optimal) and is price efficient, budget balanced and individually rational.

The equilibrium price in capacity, carbon permit and transmission right markets will be the efficient subsidy/tax for each commodity. These prices are positive only if the corresponding constraint in Problem (Socially Optimal) is binding. These prices adjust automatically to technological and economic changes. Note that at most one of the two transmission right prices corresponding to a line can be positive and it is in the direction that the line is congested; the GenCos that hold negative transmission right in the direction of congestion are therefore collecting money.

2.7.1.2 Energy-Only Design

We propose an alternative market design that uses operation reserve rather than capacity market to implement reliable capacity. In this approach, a demand profile $\tilde{U}_t(d_t, w_t)$, above the actual demand profile $U_t(d_t, w_t)$ for all $t \in T$ and $w \in W$, is auctioned; the difference between the auctioned demand and actual demand is put in operation reserve. There are several operation reserve demand profiles that can implement the reliability targets: for any large-enough operation reserve demand, the GenCos have enough incentive to invest on large amounts of capacity and therefore, the reliability target is met; but this comes at the cost of buying the large amount of operation reserve from GenCos at high price (high subsidy). Therefore, we use the results of Appendix D for long-term interaction of GenCos with individual constraints to find a method for determining an *efficient demand profile* $\tilde{U}_t(d_t, w_t)$ which can implement reliability constraints with the minimum subsidy required for energy-only approach. We determine this profile using a reverse optimization problem. Consider $\sigma^{SO}_{n,t} \in \Sigma_{n,t}, \gamma^{SO}_{n,t} \in \Gamma_{n,t} : n \in N, t \in T$ to be the solution of Problem (Socially Optimal);

choose $\tilde{U}_t^*(d_t, w_t)$ such that $\tilde{U}_t^*(d_t, w_t) \geq U_t(d_t, w_t)$, $\forall t \in T, w_t \in W_t$, and if the reliability constraint, Eq. (2.18), is omitted in Problem (Socially Optimal) and $U_t(d_t, w_t)$ is replaced by $\tilde{U}_t(d_t, w_t)$, $\sigma_{n,t}^{SO} \in \Sigma_{n,t} : n \in N, t \in T$, is still the optimal expansion strategy of the new problem, that is

$$(\sigma_{n,t}^{SO}, \gamma_{n,t} : n \in N, t \in T) = \qquad \text{(Socially Optimal - Operation Reserve)}$$

$$\arg \max_{\sigma_{n,t} \in \Sigma_{n,t}, \gamma_{n,t} \in \Gamma_{n,t} : n \in N, t \in T} E_{\mathbf{w}_{1:T}} \Bigg[\sum_{t \in T} \beta^t \tilde{U}_t(\sum_{n \in N} e_{n,t}, \mathbf{w}_t) -$$

$$\sum_{t \in T} \sum_{n \in N} \beta^t \left\{ C_{n,t}^e(e_{n,t}, \mathbf{w}_t) + C_{n,t}^x(\Delta x_{n,t}, \mathbf{w}_{t-1}) \right\} + \beta^{T+1} \sum_{n \in N} \eta(\mathbf{w}_T) x_{n,T} \Bigg] \tag{2.86}$$

$$s.t. \quad \text{Eqs. (2.16)-(2.17) and Eqs. (2.19)-(2.22)} \forall t \in T, n \in N, u, v \in V, w_{1:T} \in W_{1:T}. \tag{2.87}$$

Problem (Socially Optimal - Operation Reserve) can be used to design an energy-only solution with the following markets in each year: an elastic spot market which auctions elastic demand $\tilde{U}_t^*(\sum_{n \in N} e_{n,t}, w_t)$, an inelastic carbon permit market, and $2 \times V$ inelastic transmission right markets. The markets are designed according to the efficient auctions proposed in Section (2.3).

The proposed energy-only set of markets is different from the energy-and-capacity set of markets proposed in Section (2.7.1.1) in two ways: first, it does not have capacity markets; and second, it auctions operation reserve demand $\tilde{U}_t^*(d_t, w_t)$ instead of the real demand $U_t(d_t, w_t)$ in the spot market. In addition to these differences, Example 1 below shows that for the same level of reliable capacity, the energy-only solution requires more subsidy compared to energy-and-capacity solutions. The reason is the saturation property (Theorem 6) that holds for the energy-only solution but does not hold for energy-and-capacity solution. As a result in the energy-only solution, all of the reliable capacity should be generating power (saturated) at least at one time instant and the system operator should pay for that. But in the capacity-and-energy solution this is not necessarily true.

Example 1 (FINANCIAL COMPARISON)**.** *Consider an industry modeled in Section 2.5 where $T = 1$, i.e. single expansion decision at time 1 is followed by a single spot market at time 1^+. We model a situation where the reliable level of capacity $\underline{x}$, the cost of expansion, $C_n^x(.)$ and the cost of production $C_n^e(.)$ are high but the utility of demand, $U(.)$, is low. Particularly, assume $\underline{x} = 100MW$, there are 4 GenCos with $C_n^x(\Delta x_n) = \Delta x_n^2$, $C_n^e(e_n) = e_n^2$ and $U(d) = 10d$. Assume further that the initial capacity $x_{0,n}$ is 0 for all $n = 1, 2, 3, 4$, and that there is no retirement payment after time $T = 1$, i.e $\eta = 0\$/MW$.*

First consider the energy and capacity solutions. Since GenCos are similar in their costs and initial capacity, to meet the $100MW$ capacity requirement in a socially optimal way, each GenCo should expand $25MW$ capacity. This means the price for capacity in the capacity market is $50\$/MW$, the marginal cost of expansion by GenCos at $25MW$ capacity. Therefore, total payment by the system operator to GenCos in the capacity market is $100MW \times 50\$/MW = 5k\$/MW$. This amount *is the subsidy for reliable capacity. Moreover in the spot market, the price of electricity is marginal utility of demand which is* $10\$/MW$. *Therefore, each GenCo produces* $5MW$ *of electricity to have marginal cost of production equal to* $10\$/MW$ *and the total payment to GenCos for producing electricity at the spot market is* $4 \times 5MW \times 10\$/MW = 200\$$.

Next, consider an energy-only solution. The operation reserve demand should be set high enough to ensure GenCos have incentive to expand $25MW$ *each,* $\Delta x_n = 25MW, \quad n = 1,2,3,4$. *If they expand* $25MW$ *they should use all of it in the spot market i.e.* $e_n = 25MW, \quad n = 1,2,3,4$, *since otherwise they do not have incentive to expand that much (saturation property). Therefore the price of electricity in the spot market should be the marginal of production and expansion together equal to* $2 \times \Delta x_n + 2 \times e_n = 2 \times 25 + 2 \times 25 = 100\$/MW$. *The total payment to the GenCos is* $100\$/MW \times 100MW = 10k\$$. *To calculate the subsidy paid for reliable capacity in energy-only solution, one should calculate the payment for operation reserve; this is the difference between the total payment in the spot market,* $10000\$$, *and the efficient payment for real demand* $200\$$. *Therefore, the subsidy for reliable capacity in the energy-only solution is* $10000\$ - 200\$ = 9800\$$ *which is almost twice as large as the* $5k\$$ *efficient subsidy in the capacity-and-energy solution.*

The example above shows that energy-only solution is more expensive for the system operator because in this approach GenCos should be paid for the reliable capacity at a price equal to marginal cost of expansion plus marginal cost of production (producing at saturated capacity), $2 \times 25\$/MW + 2 \times 25\$/MW = 100\$/MW$. This is while the capacity-and-energy solution pays GenCos for the reliable capacity only at marginal cost of expansion , $2 \times 25\$/MW = 50\$/MW$.

2.7.2 Contribution to the Electricity Policy Debates

The two market designs we provided in Sections 2.7.1.1 and 2.7.1.2 have the following implications with respect to the electricity policy targets and the debates involved.

First, the design achieves price efficiency along with features (F1)-(F3) without using price caps (or offer caps), market monitoring or other sorts of market interventions. Second, with respect to reliability, our results show that no capacity markets (or other subsidiary mechanisms) are required so as to achieve socially optimal investment when planning reserve requirements do not exist; this is in contrast to References Garcia and Shen (2010) and Murphy and Smeers (2005). When planning reserve requirements exist, we proposed both an energy-and-capacity mechanism and argued that the energy-only mechanism requires more reliability subsidy to be paid to the GenCos. However, determining the demand profile required for the energy-only mechanism requires a lot of information and complex computations by the system operator; on the other hand, the energy-and-capacity mechanism requires extra markets (the capacity markets) and corresponding regulations. Third, with respect to the debate on carbon market vs. carbon tax, our design shows that the price of carbon found in the carbon permit market is the efficient level of carbon charging in order to meet the carbon emission target; in other words, the system operator can determine efficient carbon charging through the efficient carbon permit market without knowing the GenCos' private information; of course, this comes at a higher regulatory cost of running the carbon market. Finally, our transmission right market is an efficient mechanism for allocation of transmission rights proposed in References Ehrenmann and Neuhoff (2009), Chao and Peck (1996) and Hogan (1992).

Our results also contribute to the debate on charging carbon vs. incentivizing the renewables for restricting carbon emission. We briefly discuss this issue here and present the potential contribution of our results. As discussed in Section (2.1.1), policies for supporting carbon can be categorized into those that penalize carbon emission by charging carbon (through direct carbon tax or through carbon permit market) and those that reward renewables by subsidies. Those who defend the second approach argue that renewables have close to zero carbon emission factor. Furthermore, since renewables have lower marginal cost of production, they are dispatched first upon availability. As a result, increasing investment on renewables lowers carbon emission. There are downsides to supporting renewables: first, it is an indirect approach that requires determining the level of renewable capacity so as to meet the carbon emission target; such a determination is a difficult task. Second, supporting renewables has undesirable side effects such as reduced investment on conventional sources of electricity which are important for reliability (wind and solar energy are unpredictable sources of energy). Third, categorization of technologies into renewable and conventional is not sufficient to address the carbon emission problem. In reality, there are different technologies with different carbon factors. Therefore, implementing the socially optimal investment so as to restrict carbon emission can not be achieved by adding a single supporting mechanism for renewables, rather it requires supporting different technologies with different mechanisms. Finally, under certain conditions, increase in the capacity of renewables can have opposite effect i.e. increase the carbon emission. Our results for auctions with divisible goods can be used to design efficient renewable capacity markets. Such markets discover efficient subsidy for renewable capacity. Paying subsidy to renewables increases the capacity and therefore lowers the price of electricity in the spot market.

2.8 Conclusion and Future Directions

We studied efficient implementation of the three electricity policy targets, reliability, sustainability and price efficiency (together and in a uniform framework) in the restructured electricity industry where incentives and market competition are the main tools. We modeled the industry, policy targets and desired features of the markets including individual rationality, budget balanced, and social optimality. Using the model, we designed efficient electricity auctions that achieve our targets. In order to do so, we developed a framework for auctions with constraints by studying a sequence of simpler problems. We first designed efficient auctions for the fundamental problem of allocating a divisible good (elastic or inelastic good) among an oligopoly of agents with market power. Next, we used these efficient auctions to allocate divisible goods among agents with individual constraints. To allocate a divisible goods with homogeneous joint constraints (uneconomic constraints) among agents, we proposed a decomposition technique that transforms the problem into a sequence of allocations with individual constraints; each of the problems in the sequence can then be implemented using a separate efficient auction; the price in each auction determines the efficient level of tax/subsidy required for implementing the uneconomic constraint. The results where extended to multiple periods of interaction among agents with symmetric uncertainty.

Our electricity market design contributes to some of the ongoing debates on electricity policy making. We showed that price caps or market monitoring are not required for efficient prices (affordability) in the spot market. Moreover, we proposed both energy-only and energy-and-capacity mechanisms for implementing reliability, and showed that the energy-and-capacity solution is financially more efficient (less subsidy required) for the policy maker compared to energy-only solution for implementing reliability. Finally, carbon market can determine the efficient level of carbon tax.

The model we used in this paper is limited in four major aspects presented in Remark 2.2.1. Extensions for a better model of the real world are the future directions of this research. We present these extensions and discuss their implications on our results.

In contrast to our model, building new capacities takes multiple years. Considering a d-year construction period does not change the nature of the results of this paper; however, the information state for Problem (Socially Optimal - Longterm Uncertainty) and for GenCos' strategies, introduced in Lemma (3), will change to $S_t = \{x_{t-1}, \Delta x_{t-d:t-1}, w_{1:t-1}\}$ and $S_{t^+} = \{x_t, \Delta x_{t-d:t-1}, w_{1:t}\}$.

The first major future direction for this research is providing a more complete network model. We assumed there is a single demand at node 1. Distributed demand at multiple nodes requires multiple prices for electricity (nodal prices). These prices can be discovered by local markets at each node. Furthermore, if a GenCo owns owns multiple utilities in different locations, it can potentially manipulate the network by creating local congestions in the spot market. On top of that, each GenCo should decide where to expand its capacity; this extra decision factor results in a much more complicated competition among GenCos. More importantly, modeling loop flow effect due to Kirchhoff's electricity laws couples all flows in the network and therefore can change the nature of the market; in the new setup each GenCo's decision results in an externality on the other GenCos' utilities.

The second major direction for future research is considering information asymmetry among the GenCos. GenCos' private information requires implementation in other equilibrium solution concepts such as Bayesian NE or dominant strategies equilibrium. Information asymmetry results in information rent and the decrease in performance.

The third future research direction is modeling transmission expansion planning of the transmission companies (TransCos). GenCos' generation expansion problem studied in this paper is coupled to the TransCos' transmission expansion problem. For example, the high congestion price of a line is a signal for expanding the transmission capacity of that line.

CHAPTER 3

Markets with Complex Physical Constraints: Networked Electricity Spot Market

3.1 Introduction

In this chapter, we study electricity spot markets. We add Kirchhoff's laws to the model used in previous Chapter and develop a framework for networked markets using ideas from local public goods.

As discussed in the previous chapter, allocating optimal dispatch of electricity cannot be achieved by free markets and requires market design/regulation for the following features: 1) The competition is not perfect; there is an oligopoly of producers with market power. 2) The cost of producing electricity is not convex due to the fixed startup cost and economy of scale (potentially no prices exist for market clearing). 3) Any mismatch of supply and demand can cascade into blackouts with very costly consequences; this feature is a major complex constraint. 4) Electricity is not a homogeneous good across time and space because it cannot be stored.

In addition, Kirchhoff's laws introduce a loop flow effect which, combined with the lines' thermal capacity constraints, couple the producers' dispatch and limit the set of feasible dispatches to a non-convex set. Moreover, the network interconnections among producers result in externalities both negative (dispatch in the same direction) and positive (dispatch in the opposite directions).

The question of interest is how to design markets for efficient dispatch of electricity, considering all the above challenges. In the design of these markets, we are looking to incorporate prices in an effective way while still maintaining enough direct control to ensure that complex constraints are satisfied. We aim at designing markets that are 1) budget balanced, 2) individually rational, 3) price efficient, and 4) social welfare maximizing. A practical market design should also consider the information asymmetry among electricity producers which is a future direction of this research.

3.1.1 Existing Approaches for Electricity Markets with Networks

Currently, there are two approaches to electricity market design with network constraints which use two different sets of prices Ehrenmann and Neuhoff (2009): integrated market design that uses nodal prices (one price for each node in the network); and coordinated market design that uses prices for allocating transmission rights (one price for transmission right between any two nodes) and prices for allocating electricity dispatch (one price for each node in the network). Integrated market design is proposed in the form of physical transmission right (PTR), financial transmission right (FTR) and flow gate right (FGR)(Kristiansen, 2004).

The most common practice of integrated market design (nodal pricing) is supply-function auctions where producers bid their entire cost function to the system operator. The system operator then calculates the most economically efficient dispatch subject to network constraints. The price paid to each producer is the marginal cost of production at the node it is located in. Cournot competition is used as an approximation for the study of supply function auctions.

For coordinated market design, first the transmission rights and next the electricity dispatches are allocated. Transmission rights are allocated to all producers simultaneously using a pooling auction; this is because transmission rights cannot be traded bilaterally and independently due to the loop flow effect that connects all transmission rights. In PTR, the physical capacity of a line is allocated to generation companies to use it for sending electricity to other nodes in the network. In FTR only the financial right of the lines is allocated; the owner is paid for his amount of FTR at the difference in the price of electricity at the two ends of the line. Similar to FTR, FGR does not allocate physical rights, but the price paid to the holder is the marginal improvement in social welfare by increasing the capacity of the line. FGR is the correct signal for investment on increasing capacity of a line.

In competitive markets with full information, both the integrated market design and the coordinated market design lead to welfare maximizing outcome (See Bohn et al. (1984) & Cardell et al. (1997) for integrated, and Chao and Peck (1996) for coordinated). When market power exists (even under full information) the current design of these markets is not efficient Wilson (2008) and the comparison between the performance of the two approaches is not completely known Ehrenmann and Neuhoff (2009). The problem becomes even more challenging when a producer owns multiple electricity generation firms located at different nodes, or when there are multiple producers located at a single node Cardell et al. (1997). In the same way, the comparison between different approaches to coordinated markets is not clear (Joskow and Tirole, 2000). Holders of the physical right can forbid using them and this way increase the price difference at the two ends of the line, while holders of the financial right have incentive for under-producing electricity at the congested nodes so as to increase price in those nodes and collect more payment. Rules such as use-it-or-loose-it or use-it-or-sell-it are introduced for reducing such market manipulations.

The above discussion shows the necessity of designing efficient electricity spot markets with network constraints which is the question of this Chapter.

3.1.2 Contribution

We first model the electricity network and then discuss the various challenges that arise within the context of the model such as loss, lines capacity constraints and loop flow effect. Especially, to address the loop flow effect, we present the electricity network as a public good which producers and consumers should agree on how to use it. We show that with the DC approximation (first order approximation) of the flows, the set of feasible productions is convex. However, the DC approximation does not consider the loss in the lines and the set of productions with higher order approximations of the flows is non-convex. On the other hand, we show that with second order approximation of the flows, the set of feasible angles (phases) of the nodes, which uniquely determine the flow of power in the network, is convex. Angles have also the advantage that they have only local effects (the angle at each node only matters for the feasible set of angles directly connected to that).

Considering either productions (with first order approximation) or angles (with second order approximation), the problem is how to deign a market for allocation of a public good. We present such a mechanism that solves the problem. We present our mechanism by considering the angles and at the end, we compare the payments resulting from our model and mechanism with the nodal pricing payments.

3.2 Model

Consider N strategic agents in an electricity network of K nodes. Each agent is both producer and consumer. Agent $n \in N$ has maximum production capacity of x_n and his electricity production/consumption is denoted by e_n his utility being $u_n(e_n)$. The set of agents at node k is shown by N_k. Assume there are M utility companies which own the producers. The set of producers owned by utility company $m \in N$ is denoted by N^m. The operating angle of node k is θ_k. The flow of power from node k to node j is denoted by I_{kj} determined by a function $f_{kj}\left(\theta_k - \theta_j\right)$. The thermal capacity of the line between nodes k and j is denoted by L_{kj}. Agents know each other's utilities but the the system operator does not know the utilities. All other information is common knowledge among producers and the system operator.

The socially optimal (centralized) electricity dispatch problem is:

$$\max_{\theta_k, e_n, k\in K, n\in N} \sum_{n\in N} u_n(e_n) \tag{3.1}$$

$$s.t. \quad 0 \le e_n \le x_n \quad \forall n \in N \tag{3.2}$$

$$\sum_{n\in N_k} e_n = \sum_{j\in K} I_{kj} \quad \forall k \in K \tag{3.3}$$

$$I_{kj} \le L_{kj} \quad \forall k, j \in K \tag{3.4}$$

$$I_{kj} = f_{kj}(\theta_k - \theta_j) \quad \forall k, j \in K \tag{3.5}$$

In reality, the function f has a complex sinusoid form. Therefore, we approximate it. The first-order approximation is a linear one, known as DC approximation; Consider B_{kj} to be the first derivative of f_{kj} at zero. Then the DC approximation is described by

$$I_{kj} = f_{kj}(\theta_k - \theta_j) = B_{kj} \times (\theta_k - \theta_j) \tag{3.6}$$

The second approximation is a second order polynomial. Consider G_{kj} to be the second derivative of f_{kj} at zero.

$$I_{kj} = B_{kj} \times (\theta_k - \theta_j) + \frac{1}{2} G_{kj} \times (\theta_k - \theta_j)^2 \tag{3.7}$$

In this approximation the second order term captures the loss of power along the line. The total loss in line kj is $I_{kj} + I_{jk} = G_{kj} \times (\theta_k - \theta_j)^2$.

We wish to design auctions to implement the solution of the above problem. We would like these auctions to be budget balanced, individually rational and price efficient. Price efficiency in this setup means that each agent pays his marginal utility with respect to each angle as his price for that angle.

3.3 Elements of Networked Markets

The model in Section 3.2 departs from the model in the previous Chapter (where several strategic producers and a single non-strategic demand trade electricity at the same node) in the following ways:

1. The lines have limited capacity. The capacity of a line limits the trade in between its two ends. If a line is congested, then the markets on the two sides of the line are not connected, instead they are two separate local markets. The producers located on each side have local power and the markets have separate prices.

2. The flow of electricity is subject to loop flow constraints (KVL and KCL). These constraints make electricity networks unique and different from any other network; they have complicated effects as they connect all flows in the network to each other. For example, there could be the case that the demand at two nodes which are connected to each other with a non-congested line pay different prices for electricity. As a result of the connectivity of productions due to the Kirchhoff's laws, a producer and a consumer cannot do a separate bilateral trade independent of the others. Instead, all trades should be cleared at the same time and in a pooling market. Also, network becomes a public good whose use should be determined by all the producers and consumers together.

3. The lines have loss in delivery of electricity power. This loss is a function of the total power flow over the line. The power delivered at the one end of the line is the total power injected at the opposite end minus this loss. In markets, this loss should be shared between the buyer and the seller and this is reflected in the electricity prices. For example, consider a simple network of one producer and one consumer connected with a lossy line. If the trading commodity is electricity at the producer's node, then the consumer should endogenize the cost of loss in his bids. On the other hand, if the trading commodity is electricity at the consumer's node then the producer should endogenize the cost of loss in her bids. Alternatively, the producer and consumer can trade electricity at some point in the line between them. In this case they do loss sharing. Based on this simple example, one can observe that the social welfare maximizing solution is the same for all the cases but the payments of the producers and consumer dependon the location where the trade takes place.

4. Multiple producers are located at the same node. As a result, even if the system operator determines the outflow of electricity power from every node, each agent's share in that node should be determined. This may require local markets among the agents at each node.

5. Multiple producers at different nodes can be owned by a single generation company. This single company then maximizes its aggregate utility from all the producers it owns. As a result, it may strategically produce more at one of its producers (have negative utility for that single producer) and congest the lines. This way, it can decompose/break the network into separate local markets and then exercise its local market power through its producers at each separate local market.

6. Demand consists of multiple consumers at different nodes. Multiple consumers can be separated by the network constraints. Therefore, they may each have a separate market clearing price for buying electricity. This necessitates multiple markets for discovering these separate prices.

7. Demand can be strategic. Strategic demand is a major departure from the model in the previous chapter. Strategic demand is realistic with smart grids, where the demand is price sensitive and price responsive. Strategic demand can bid in the market.

3.4 Market Design for Electricity Trade over Network

We consider a model that includes features $1,2,3,6,7$ presented in Section 3.3, that is, the model includes strategic demand, maximum of one agent at each node and second order approximation of the flows (which considers losses). For this model, we present a local public good market that implements socially optimal dispatch of electricity. In this market, the nodes' angles are the public good. For the DC approximation (no loss), the set of feasible production profiles of electricity is convex; in this case, we can design a market where the electricity production profiles are the public good.

Consider a variation of the model in Section (3.2) where at each node, there is exactly one strategic agent (producer or consumer). Moreover, each node is connected to at least two other nodes (this assumption is required for our market design). Define N to be the set of all nodes and R_n as the set of nodes that are connected to node n including node n itself. θ_{R_n} denotes the set of angles of these nodes (note that $|\theta_{R_n}| \geq 3$).

Considering the flow from node n to node k as

$$I_{nk} = f_{nk}(\theta_n - \theta_k) = B_{nk} \times (\theta_n - \theta_k) + \frac{1}{2} G_{nk} \times (\theta_n - \theta_k)^2, \tag{3.8}$$

the production/consumption at node n is

$$e_n(\theta_{R_n}) = \sum_{k \in R_n} B_{nk} \times (\theta_n - \theta_k) + \frac{1}{2} G_{nk} \times (\theta_n - \theta_k)^2 \tag{3.9}$$

Note that both I_{nk} and e_{nk} are convex functions of θ_{R_n}. If $e_n > 0$, then node n is a producer, otherwise it is a consumer. The utility for producing/consuming electricity at node n is $u_n(e_n(\theta_{R_n}))$. Assume u_n is a strictly concave and decreasing function of e_n with $u_n(0) = 0$. Since e_n is a convex function of θ_{R_n}, u_n is also a strictly concave function of θ_{R_n} (See Rockefeller Theorem 5.1).

3.4.1 The Centralized Problem and The Associated KKT Conditions

The centralized electricity dispatch problem is:

$$\theta_N^* = \arg\max_{\theta_n, n\in N} \sum_{n\in N} u_n(\theta_{R_n}) \qquad \text{(Socially Optimal)}$$

$$s.t. \quad e_n(\theta_{R_n}) \leq x_n \quad \forall n \in N \tag{3.10}$$

$$I_{nk} \leq L_{nk} \quad \forall n,k \in N \tag{3.11}$$

where I_{nk} and e_n are determined by Equations (3.8) and (3.9).

This maximization has a strictly concave objective function. The domain is convex because I_{kj} is a convex function of θs and therefore the set of θs satisfying $I_{kj} \leq L_{kj}$ is convex, and furthermore, the intersection of a finite number of convex sets is also convex. Define this domain by $Dom(\theta_N)$. As a result, the above optimization is a strictly concave optimization. We present the Lagrangian and KKT conditions for this problem. The Lagrangian is

$$L = \sum_{n\in N} u_n(\theta_{R_n}) + \sum_{n\in N} \mu_n(x_n - e_n(\theta_{R_n})) + \sum_{n,k\in N} \lambda_{nk}(L_{nk} - I_{nk}) \tag{3.12}$$

Considering that

$$\frac{\partial I_{nk}}{\partial \theta_k} = -B_{nk} - G_{nk} \times (\theta_n - \theta_k) \tag{3.13}$$

$$\frac{\partial I_{nk}}{\partial \theta_n} = B_{nk} + G_{nk} \times (\theta_n - \theta_k) \tag{3.14}$$

and

$$\frac{\partial e_n(\theta_{R_n})}{\partial \theta_k} = \begin{cases} -B_{nk} - G_{nk} \times (\theta_n - \theta_k) & k \neq n, \\ \sum_{i\in R_n} B_{ni} + G_{ni} \times (\theta_n - \theta_i), & k = n, \end{cases} \tag{3.15}$$

the KKT conditions are

$$\frac{\partial L}{\partial \theta_n}|_{\theta^*_N,\lambda^*_{N\times N},\mu^*_N} = [\sum_{k\in R_n}[u'_k(e_k)\frac{\partial e_k(\theta_{R_k})}{\partial \theta_n} - \mu_k\frac{\partial e_k}{\partial \theta_n} - \lambda_{nk}\frac{\partial I_{nk}}{\partial \theta_n} - \lambda_{kn}\frac{\partial I_{kn}}{\partial \theta_n}]|_{\theta^*_N,\lambda^*_N,\mu^*_N} =$$

$$\Big[u'_n(e_n)\times[\sum_{i\in R_n} B_{ni}+G_{ni}\times(\theta_n-\theta_i)] + \sum_{k\neq n,k\in R_n} u'(e_k)\times[-B_{kn}-G_{kn}\times(\theta_k-\theta_n)]$$

$$-\mu_n\times[\sum_{i\in R_n} B_{ni}+G_{ni}\times(\theta_n-\theta_i)] - \sum_{k\in R_n,k\neq n}\mu_k[-B_{nk}-G_{nk}\times(\theta_n-\theta_k)]$$

$$-\sum_{k\in R_n}\lambda_{nk}[+B_{nk}+G_{nk}\times(\theta_n-\theta_k)] - \sum_{k\in R_n}\lambda_{kn}\times[-B_{kn}-G_{kn}(\theta_k-\theta_n)]\Big]|_{\theta^*_N,\lambda^*_{N\times N},\mu^*_N} = 0 \quad (3.16)$$

$$\lambda_{kn}(L_{kn}-I_{kn})|_{\theta^*_N,\lambda^*_{N\times N},\mu^*_N} = 0, \quad \forall k,n\in N \quad (3.17)$$

$$\mu_n(x_n-e_n)|_{\theta^*_N,\lambda^*_{N\times N},\mu^*_N} = 0, \qquad \forall n\in N \quad (3.18)$$

$$\lambda_{nk}\geq 0, \mu_n\geq 0|_{\theta^*_N,\lambda^*_{N\times N},\mu^*_N}. \quad (3.19)$$

3.4.2 The Local Public Good Mechanism

We present the following local public good mechanism that is social welfare maximizing, individually rational and price efficient. In this mechanism, the network is a public resource whose usage is decided by all the strategic agents collectively.

Message Space:

Agent located at node n sends the following message:

$$m_n = \{\hat{\theta}^n_{R_n}, p^n_{R_n}\} \quad (3.20)$$

$$\hat{\theta}_{R_n} \in Dom(\theta_{R_n}), p^n_{R_n} \in \mathbb{R}^{+|R_n|}; \quad (3.21)$$

thus, $\hat{\theta}_{R_n}$ consists of the angles proposed for all the nodes in R_n and $p^n_{R_n}$ is a set of prices proposed for these nodes.

Outcome Function

The angle of node n is set as

$$\theta_n(m_{R_n}) = \frac{1}{|R_n|}\sum_{k\in R_n}\hat{\theta}^k_n \quad (3.22)$$

Since $\hat{\theta}^k_n \in Dom(\theta_{R_n}), \forall k\in N$ and $Dom(\theta_n)$ is convex set, $\theta_n(m_{R_n})$ is also in $Dom(\theta_n)$.

Assume agents in R_k are indexed from 1 to $|R_K|$ and $y_{n,k}$ refers to the index of agent n in R_k.

The payment of gent at node n is

$$t_n(m_{R_n}) = \sum_{k\in R_n} l_{nk}(m_{R_n})\theta_k(m_{R_n}) + \sum_{k\in R_n} p_k^n(\hat{\theta}_k^n - \hat{\theta}_k^{y_{n,k}+1})^2 \tag{3.23}$$

$$l_{nk}(m_{R_n}) = p_k^{y_{n,k}+1} - p_k^{y_{n,k}+2}. \tag{3.24}$$

The total utility of agent n is

$$u_n(\theta_{R_n}) - t_n. \tag{3.25}$$

3.4.3 Properties of the Mechanism

Before we introduce the properties of the mechanism, we study the agents' best response in the game induced by the mechanism. An agent's best response will be useful in establishing the properties of the mechanism. In the game induced by this mechanism, agent n's best response to the message m_{-n} of the other agents is

$$\max_{\hat{\theta}_{R_n}^n, p_{R_n}^n} u_n\big(e_n(\theta_{R_n}(m_{-n}, \hat{\theta}_{R_n}^n))\big) - t_n(m_{-n}, \hat{\theta}_{R_n}^n, p_{R_n}^n) \tag{Best Response}$$

$$s.t. \quad e_n \le x_n \tag{3.26}$$

$$I_{nk} \le L_{nk} \quad \forall k \in R_n \tag{3.27}$$

$$I_{kn} \le L_{kn} \quad \forall k \in R_n \tag{3.28}$$

$$p_k^n \ge 0 \quad k \in R_n \tag{3.29}$$

where I_{nk} and e_n are determined by Equations (3.8) and (3.9).

This is a strictly concave optimization problem in m_n. We present the Lagrangian and the KKT conditions for the problem. The Lagrangian is

$$\begin{aligned} H_n = {} & u_n(e_n\big(\theta_{R_n}(m_{-n}, \hat{\theta}_{R_n}^n)\big)) - t_n(m_{-n}, \hat{\theta}_{R_n}^n, P_{R_n}^n) + \hat{\mu}_n(x_n - e_n) \\ & + \sum_{k\in R_n} [\hat{\lambda}_{nk}(L_{nk} - I_{nk}) + \hat{\lambda}_{kn}(L_{kn} - I_{kn})] + \sum_{k\in R_n} \hat{\gamma}_{nk} p_k^n \end{aligned} \tag{3.30}$$

Considering that

$$\frac{\partial \theta_n}{\partial \hat{\theta}_n^k} = \frac{1}{|R_n|}, \tag{3.31}$$

the KKT conditions at $m_{-n}, m_n^*, \hat{\lambda}_{nk}^*, \hat{\mu}_n^*, \hat{\gamma}_n^*$ are:

$$\frac{\partial H_n}{\partial \hat{\theta}_n^n} = u_n'(e_n) \times \frac{\partial e_n}{\partial \theta_n} \times \frac{\partial \theta_n}{\partial \hat{\theta}_n^n} - \frac{\partial t_n}{\partial \hat{\theta}_n^n}$$
$$- \hat{\mu}_n \frac{\partial e_n}{\partial \theta_n} \frac{\partial \theta_n}{\partial \hat{\theta}_n^n} - \sum_{k \in R_n} [\hat{\lambda}_{nk} \frac{\partial I_{nk}}{\partial \theta_n} \frac{\partial \theta_n}{\partial \hat{\theta}_n^n} + \hat{\lambda}_{kn} \frac{\partial I_{kn}}{\partial \theta_n} \frac{\partial \theta_n}{\partial \hat{\theta}_n^n}]$$
$$= u_n'(e_n)[\sum_{i \in R_n} B_{ni} + G_{ni} \times (\theta_n - \theta_i)] \times \frac{1}{|R_n|} - l_{nn} \frac{1}{|R_n|} - 2p_n^n(\hat{\theta}_n^n - \hat{\theta}_n^{y_{n,n}+1}) +$$
$$- \hat{\mu}_n [\sum_{i \in R_n} B_{ni} + G_{ni} \times (\theta_n - \theta_i)] \frac{1}{|R_n|}$$
$$\sum_{k \in R_n} [\hat{\lambda}_{nk}(B_{nk} + G_{nk}(\theta_n - \theta_k)) \frac{1}{|R_n|} + \hat{\lambda}_{kn}(-B_{kn} - G_{kn}(\theta_k - \theta_n)) \frac{1}{|R_n|}] = 0 \tag{3.32}$$
$$\frac{\partial H_n}{\partial \hat{\theta}_k^n} = u_n'(e_n) \times \frac{\partial e_n}{\partial \theta_k} \times \frac{\partial \theta_k}{\partial \hat{\theta}_k^n} - \frac{\partial t_n}{\partial \hat{\theta}_k^n}$$
$$- \hat{\mu}_n \frac{\partial e_n}{\partial \theta_k} \frac{\partial \theta_k}{\partial \hat{\theta}_n^k} - \hat{\lambda}_{nk} \frac{\partial I_{nk}}{\partial \hat{\theta}_k^n} - \hat{\lambda}_{kn} \frac{\partial I_{kn}}{\partial \hat{\theta}_k^n} =$$
$$u_n'(e_n)[-B_{nk} - G_{nk} \times (\theta_n - \theta_k)] \times \frac{1}{|R_k|} - l_{nk} \frac{1}{|R_k|} - 2p_k^n(\hat{\theta}_k^n - \hat{\theta}_k^{y_{n,k}+1}) +$$
$$- \hat{\mu}_n [-B_{nk} - G_{nk} \times (\theta_n - \theta_k)] \frac{1}{|R_k|}$$
$$- \hat{\lambda}_{nk}(B_{nk} + G_{nk}(\theta_n - \theta_k)) \frac{1}{|R_k|} - \hat{\lambda}_{kn}(-B_{kn} - G_{kn}(\theta_k - \theta_n)) \frac{1}{|R_k|} \quad \forall k \in R_n, k \neq n \tag{3.33}$$
$$\frac{\partial H_n}{\partial p_k^n} = -\frac{\partial t_n}{\partial p_k^n} = (\hat{\theta}_k^n - \hat{\theta}_k^{y_{n,1}})^2 + \hat{\gamma}_{nk} = 0 \quad \forall k \in R_n \tag{3.34}$$
$$\hat{\lambda}_{nk}(L_{nk} - I_{nk}) = 0 \tag{3.35}$$
$$\hat{\mu}_n(x_n - e_n) = 0, \quad \hat{\gamma}_{nk} p_k^n \geq 0 \tag{3.36}$$
$$\hat{\lambda}_{nk}, \hat{\mu}_n, \hat{\gamma}_n \geq 0 \tag{3.37}$$

Using the above mechanism, we prove the proposed mechanism possesses the following properties.

Theorem 11. *The set of Nash Equilibria of the game induced by the mechanism is non-empty. Furthermore, the outcome corresponding to each Nash equilibrium (NE) is a solution of the centralized problem of Section 3.4.1, and at each NE the mechanism is budget-balanced, individually rational and price efficient.*

Every Nash Equilibrium of the game induced by the mechanism is budget balanced, socially optimal, individually rational, and price efficient. Moreover, the set of NE of the game is non-empty.

Proof. Let m_N^* be a Nash Equilibrium of the game induced by the mechanism. We establish the assertion of this theorem in the following steps.

Step 1: We first prove that at equilibrium,

$$t_n^* = \sum_{k \in N} l_{nk}^* \theta_k^* \tag{3.38}$$

or equivalently,

$$p_k^n (\hat{\theta}_k^n - \hat{\theta}_k^{y_{n,k}+1})^2 = 0 \quad \forall k \in R_n \tag{3.39}$$

We establish Eq. (3.39) by contradiction. Assume $p_k^n (\hat{\theta}_k^n - \hat{\theta}_k^{y_{n,k}+1})^2 \neq 0$; then agent n can change his price proposal to $p_k^n = 0$, without changing the rest of his message, and increase his utility. This means m_n^* is not his best response to m_{-n}^*, a contradiction.

Step 2: By construction $\sum_{k \in R_n} l_{nk} = 0$. Therefore, using the first step, the total payments in the system add up to zero:

$$\sum_{n \in N} t_n^* = \sum_{n \in N} (\sum_{k \in R_n} l_{nk}) \times \theta_n = 0. \tag{3.40}$$

This means the mechanism is budget balanced at equilibrium.

Step 3: To prove that the outcome corresponding to every NE of the game induced by the mechanism is socially optimal, we construct the KKT conditions of Problem (Socially Optimal), from the KKT conditions of Problem (Best Response). Let $\hat{\theta}_{R_n}^{n^*}, {p_{R_n}^n}^*, \hat{\lambda}_{nk}^*, \hat{\mu}_n^*, \hat{\gamma}_{nk}^*, n, k \in N$ describe a NE and the corresponding dual variables. We claim that the following set of variables is a solution for Problem (Socially Optimal).

$$\theta_n = \frac{\sum_{k \in R_n} \hat{\theta}_n^{k^*}}{|R_n|} \tag{3.41}$$

$$\lambda_{nk} = 2 \times \hat{\lambda}_{nk}^*, n \neq k \tag{3.42}$$

$$\mu_n = \hat{\mu}_n^* \tag{3.43}$$

To prove this claim, we construct the KKT conditions of Problem (Socially Optimal). Note that by

Eq. (3.32) and (3.33)

$$\sum_{k\in R_n}\frac{\partial L_k}{\partial\hat{\theta}_n^k}=[u_n'(e_n)\frac{\partial e_n(\theta_{R_n})}{\partial\theta_n}-\hat{\mu}_n\frac{\partial e_n}{\partial\theta_n}+\sum_{k\in R_n,k\neq n}-\hat{\lambda}_{nk}\frac{\partial I_{nk}}{\partial\theta_n}-\hat{\lambda}_{kn}\frac{\partial I_{kn}}{\partial\theta_n}]\times\frac{1}{|R_n|}$$
$$+\sum_{k\in R_n,k\neq n}[u_k'(e_k)\frac{\partial e_k(\theta_{R_k})}{\partial\theta_n}-\hat{\mu}_k\frac{\partial e_k}{\partial\theta_n}+$$
$$-\hat{\lambda}_{nk}\frac{\partial I_{nk}}{\partial\theta_n}-\hat{\lambda}_{kn}\frac{\partial I_{kn}}{\partial\theta_n}]\times\frac{1}{|R_n|}+\sum_{k\in R_n}l_{nk}\frac{1}{|R_n|}+\sum_{k\in R_n}-2p_n^k(\hat{\theta}_n^k-\hat{\theta}_n^{y_{k,n}+1})=0 \tag{3.44}$$

Considering Eq. (3.44) and Eq. (3.39), along with the fact that by construction $\sum_{k\in R_n} l_{nk}=0$, and taking into account the variables defined in Eqs. (3.41)-(3.43) we obtain.

$$u_k'(e_k)\frac{\partial e_k(\theta_{R_k})}{\partial\theta_n}-\mu_k\frac{\partial e_k}{\partial\theta_n}-\lambda_{nk}\frac{\partial I_{nk}}{\partial\theta_n}-\lambda_{kn}\frac{\partial I_{kn}}{\partial\theta_n}=0, \tag{3.45}$$

which is Eq. (3.16), the first KKT condition of Problem (Socially Optimal). The other KKT conditions, Eqs. (3.17)-(3.19), hold because of Eqs. (3.34)-(3.56).

Step 4: We prove the existence of NE for the game induced by the mechanism in two stages. In the first stage, we introduce a new optimization problem for each agent, called surrogate optimization problem. We show that collectively, these surrogate optimization problems have the same solution as the centralized problem (Socially Optimal). Second, we use the surrogate optimization problems to prove the existence of NE for the game induced by the mechanism.

First Stage: Let $(\theta_N^*,\lambda_N^*,\mu_N^*)$ denote the unique solution and corresponding dual variables of Problem (Socially Optimal). Define the following individual prices calculated at $(\theta_N^*,\lambda_N^*,\mu_N^*)$.

$$\tilde{l}_{n,n}^*=u_n'(e_n)\times\frac{\partial e_n}{\partial\theta_n}-\mu_n\frac{\partial e_n}{\partial\theta_n}-\sum_{k\in R_n}[\frac{\lambda_{nk}}{2}\frac{\partial I_{nk}}{\partial\theta_n}+\frac{\lambda_{kn}}{2}\frac{\partial I_{kn}}{\partial\theta_n}] \tag{3.46}$$
$$\tilde{l}_{n,k}^*=u_n'(e_n)\times\frac{\partial e_n}{\partial\theta_k}-\mu_n\frac{\partial e_n}{\partial\theta_k}-\frac{\lambda_{nk}}{2}\frac{\partial I_{nk}}{\partial\theta_k}-\frac{\lambda_{kn}}{2}\frac{\partial I_{kn}}{\partial\theta_k}. \tag{3.47}$$

Consider the following (individual) optimization problem for n.

$$\max_{\theta_{R_n}^{n,sur}} u_n\big(e_n(\theta_{R_n})\big)-\sum_{k\in R_n}\tilde{l}_{nk}^*\theta_k \tag{Surrogate Optimization}$$
$$s.t.\quad e_n\leq x_n \tag{3.48}$$
$$I_{nk}\leq L_{nk}\quad\forall k\in R_n \tag{3.49}$$
$$I_{kn}\leq L_{kn}\quad\forall k\in R_n \tag{3.50}$$

where I_{nk} and e_n are determined by Equations (3.8) and (3.9). This is a strictly concave optimization in $\theta^n_{R_n}$. We present the Lagrangian and the KKT conditions for the problem. The Lagrangian is

$$H_n^{sur} = u_n\big(e_n\big(\theta^n_{R_n}\big)\big) - \sum_{k\in R_n} \tilde{l}^*_{nk}\theta_k + \mu_n^{sur}(x_n - e_n)$$
$$+ \sum_{k\in R_n} [\lambda_{nk}^{sur}(L_{nk} - I_{nk}) + \lambda_{kn}^{sur}(L_{kn} - I_{kn})] \tag{3.51}$$

The KKT conditions at $\theta^{n*}_{R_n}, \lambda^{sur*}_{nk}, \mu^{sur*}_n$ are:

$$\frac{\partial H_n^{sur}}{\partial \theta_n^{n,sur}} = u_n'(e_n)\times\frac{\partial e_n}{\partial \theta_n^{n,sur}} - \tilde{l}^*_{nn} - \hat{\mu}_n\frac{\partial e_n}{\partial \theta_n^{n,sur}} - \sum_{k\in R_n}[\lambda_{nk}^{sur}\frac{\partial I_{nk}}{\partial \theta_n^{n,sur}} + \lambda_{kn}^{sur}\frac{\partial I_{kn}}{\partial \theta_n^{n,sur}}] \tag{3.52}$$

$$\frac{\partial H_n^{sur}}{\partial \theta_k^{n,sur}} = u_n'(e_n)\times\frac{\partial e_n}{\partial \theta_k^{n,sur}} - \tilde{l}^*_{nk} - \mu_n^{sur}\frac{\partial e_n}{\partial \theta_k^{n,sur}} - \lambda_{nk}^{sur}\frac{\partial I_{nk}}{\partial \theta_k^{n,sur}} - \lambda_{kn}^{sur}\frac{\partial I_{kn}}{\partial \theta_k^{n,sur}} = 0 \quad \forall k\in R_n, k\neq n \tag{3.53}$$

$$\lambda_{nk}^{sur}(L_{nk} - I_{nk}) = 0 \tag{3.54}$$

$$\mu_n^{sur}(x_n - e_n) = 0, \tag{3.55}$$

$$\lambda_{nk}^{sur}, \mu_n^{sur}, \gamma_n^{sur} \geq 0 \tag{3.56}$$

Using Eqs. (3.46) and (3.47) it is straightforward to show that $(\theta^*_{R_n}, \mu^*_n, \frac{\lambda^*_{nk}}{2}) : k \in R_n)$ from the solution of Problem (Socially Optimal) is a solution to the surrogate optimization problem at n. Thus collectively these surrogate optimization problems result in the solution of the centralized (socially optimal) optimization problem.

Second Step: We construct a NE of the game induced by the mechanism by showing that the KKT conditions of Problem (Best Response) are satisfied. Let $r^* = (\theta^*_N, \lambda^*_{N\times N}, \mu^*_N)$ be the solution to Problem (Socially Optimal). From the first stage above, we know that this is also a solution for the surrogate problems.

Consider $\hat{r}^* = (\hat{\theta}^*_N, p^*_N, \hat{\mu}^*_N, \hat{\lambda}^*_{N\times N})$ to be a solution to the following equations:

$$\hat{\theta}^k_n = \theta^*_n \tag{3.57}$$

$$p^k_n - p_n^{y_{k,n}+1} = \tilde{l}_{kn} \tag{3.58}$$

$$p^k_n \geq 0 \tag{3.59}$$

$$\hat{\lambda}_{kn} = \lambda^*_{kn}/2 \tag{3.60}$$

$$\hat{\mu}_n = \mu^*_n \tag{3.61}$$

Note that for Eq. (3.58) to have a solution, we should have

$$\sum_{j \in R_k} \tilde{l}_{jk} = 0 \tag{3.62}$$

which is true by Eq. (3.16).

The KKT conditions of Problem (Surrogate Optimization) show that at $\hat{r}^*$, the KKT conditions for Problem (Best Response) are satisfied and therefore, $\hat{r}^*$ defines a NE and the corresponding dual variables.

Step 5: As shown in Step 3, the outcome corresponding to any NE of the game induced by the mechanism is a solution of Problem (Socially Optimal). The prices at equilibrium are given by Eqs. (3.46) and (3.47). These are the efficient prices and consist of three parts: marginal cost of production with respect to angles, $u_n'(e_n) \times \frac{\partial e_n}{\partial \theta_k}$, the saturation price for limited production capacity $-\mu_n \frac{\partial e_n}{\partial \theta_k}$, and the lines congestion price $-\frac{\lambda_{nk}}{2}\frac{\partial I_{nk}}{\partial \theta_k^n} - \frac{\lambda_{kn}}{2}\frac{\partial I_{kn}}{\partial \theta_k^n}$.

Step 6: To show that the mechanism is individually rational, consider a NE $(\hat{\theta}_k^{n*}, p_k^{n*} : n \in N, k \in N_n)$ of the game induced by the mechanism. If GenCo n deviates from his equilibrium message to

$$\hat{\theta}_k^n = -\frac{1}{|R_n|} \sum_{j \in R_n, j \neq k} \hat{\theta}_n^{j*}, \quad \forall k \in R_n \tag{3.63}$$

$$p_k^n = 0 \quad \forall k \in R_n, \tag{3.64}$$

then the outcome from this deviation will be $e_n = 0$ and $t_n = 0$ which gives agent n a utility of zero. Note that the message $(\hat{\theta}_k^n, p_k^n)$ given by Eqs. (3.63) and (3.64) is a feasible message, since $Dom(\theta_n)$ is a symmetric convex set and therefore, for $\hat{\theta}_n^j \in Dom(\theta_n)$, the negative of their average is also in $Dom(\theta_n)$. Since by definition GenCo n should not gain by deviation from the NE, his utility at equilibrium should be non-negative. □

Theorem 12. *At equilibrium, each agent's payment consists of the followings: (i) his production/consumption at a price equal to his marginal utility, (ii) half of the loss for his node at a price equal to the marginal utility, and (iii) the power flow from his node to each line going out of it minus half of the loss at the congestion cost of the link. This means the producers and consumers share the cost of loss equally.*

When there is no loss in the system (i.e. DC approximation of the flows), this payment are according to the nodal pricing system.

Proof. By Theorem 11, at the equilibrium of the game induced by the mechanism $\hat{\theta}_{R_n}^* = \theta_{R_n}^*$ and the individual prices l_{nk}^* are determined by Eqs. (3.46) and (3.47).

Define $loss_{nk}$ to be the loss in line from n to k, and $loss_n$ to be the total loss in all the lines going out of node n; then,

$$loss_{nk} = \frac{1}{2} G_{nk} (\theta_n - \theta_k)^2 \tag{3.65}$$

and

$$loss_n = \sum_{k \in N} \frac{1}{2} G_{nk} (\theta_n - \theta_k)^2 \tag{3.66}$$

The equilibrium payment to the agent located at node n, according to Eqs. (3.40), (3.46) and (3.47), is

$$t_n^* = \sum_{j \in R_n} l_{nj}^* \theta_j^* = \tag{3.67}$$

$$[u_n'(e_n^*) - \mu_n^*] \times [\sum_{j \in R_n, j \neq n} (-B_{nj} - G_{nj}(\theta_n^* - \theta_j^*)) \times \theta_j^*$$

$$+ (\sum_{j \in R_n} B_{nj} + G_{nj} \times (\theta_n^* - \theta_j^*)) \times \theta_n^*] + \sum_{j \in R_n} \frac{\lambda_{nk}^* + \lambda_{kn}^*}{2} [B_{nk}(\theta_n^* - \theta_k^*) + G_{nk}(\theta_n^* - \theta_k^*)^2] \tag{3.68}$$

$$= [u_n'(e_n^*) - \mu_n^*] \times [e_n^* + \frac{1}{2} Loss_n^*] + \sum_{j \in R_n} (\frac{\lambda_{nk}^* + \lambda_{kn}^*}{2})(I_{nk} + \frac{loss_{nk}^*}{2}) \tag{3.69}$$

This payment consists of the parts mentioned in the statement of the Theorem.

When the loss is zero, the payments are equal to those of the nodal pricing system. □

3.5 Future Directions

The mechanism designed in this chapter can be extended in the following directions. First consider the case where there exist nodes without any producer or consumer located in them or in their neighborhood. According to the mechanism designed in this section, no one bids for angles at such nodes. The definition of neighboring nodes should be extended to cover this case. Second, the model must be extended to cover the situations where there are multiple strategic agents at one node (See issue/feature number 4 in Section 3.3), as well as the case where there are non-strategic consumers (See issue number 7 in Section 3.3).

Finally, the approach used in this Section for electricity networks can be extended to other networked markets with externalities such as spectrum allocation in a wireless network with interference. According to our approach, the network can be considered as a public good. Each agent in the network will have individual prices for the network. In large-scale networks, forming a local public goods can reduce the complexity of the mechanism.

CHAPTER 4

Large-Scale CPS: Network Security with Non-Bayesian Uncertainty

4.1 Introduction

The defense of computer systems (cyber security) plays a crucial role in their efficient/normal operation. One class of cyber security problems concerns the security of networks of computers (cyber networks).[1] In this work, we investigate the development of defense policies for the security of cyber networks.

There are several approaches to addressing the cyber security problem. These approaches can be categorized into static vs. dynamic and control-theoretic vs. game-theoretic. The static approach considers a one-stage/single-period decision problem where the goal is to determine a defense policy. The dynamic approach considers a multi-period decision problem where the goal is to determine a feedback defense policy that takes into account the evolution of the system and the available information over time. Both the static and dynamic approaches can vary in the assumptions on the attacker's behavior (strategic vs. non-strategic). Strategic behavior leads to a game, *e.g.* Alpcan and Başar (2010); Lye and Wing (2005); Manshaei et al. (2013), whereas non-strategic behavior leads either to an optimization problem in the static case or a control problem in the dynamic case, *e.g.* Ligatti et al. (2005); Miehling et al. (2015); Rasouli et al. (2014); Schneider (2000). In each of the above categories, there exist various assumptions on the problem's *information structure*, that is, the information each agent possesses at each time instant. The information structure can be symmetric (both agents possess the same information) or asymmetric (agents possess differing information).

In this chapter, we approach the security of cyber networks as a control problem from the defender's point-of-view. We model the attacker as nature. The security status of the cyber network evolves over time as a function of both the defender's and nature's actions. We assume that the defender does not possess complete information of the security status of the network at

[1] These cyber networks are typically very large.

any given time. Due to the defender's lack of complete information of the security status, we take a conservative approach to determining a defense policy. Specifically, we seek to minimize the worst-case damage that the attacker/nature can inflict on the cyber network. Therefore, we determine a defense policy as the solution of a minmax control problem with imperfect information. Due to the high-dimensionality of the minmax control problem we cannot solve it precisely. As a result, we develop a scalable approach to its solution, resulting in a suboptimal/approximate solution of the original problem. The approach is based on the concept of an influence graph (which quantifies the functional dependencies between the problem's variables) and uses a clustering algorithm to decompose the original, high-dimensional minmax control problem into a collection of lower-dimensional minmax problems that are computationally feasible.

Our approach captures the dynamic nature of attacks and the fact that the defender does not possess perfect knowledge of the security status of the network. Even though we do not model the attacker as a strategic agent, we compensate (in part) for the lack of this feature by adopting a minmax performance criterion, which leads to a conservative defense approach.

Our work is distinctly different from the existing literature. From a security perspective, our work falls within the category of *intrusion response systems* (IRSs), where there is a rich literature (see Foo et al. (2008); Inayat et al. (2016); Shameli-Sendi et al. (2012) and references therein). The goal of an IRS is to take in security information from the environment and translate it into a defense action, with the goal of interfering with an attacker's objective(s). To the best of the authors' knowledge, our work is the first to investigate the design of an IRS from a minmax control perspective. From a control theory point of view, our model and problem are similar to those of Baras and James (1994); Bernhard (1995, 2000, 2003); Bertsekas and Rhodes (1973); Bertsekas (1971); Bertsekas and Rhodes (1971); Coraluppi and Marcus (1999); Witsenhausen (1968, 1966); however, one distinguishing feature of our model is in the structure of observations, allowing us to capture some essential features of cyber security problems.

4.1.1 Organization of the Chapter

The chapter is organized as follows. In Section 4.2, we introduce the security model. In Section 4.3, we formulate the defense problem, define the notion of an *information state* for the problem, and describe a sequential decomposition procedure for the problem's solution. In Section 4.4, we describe our approximation approach to the defense problem. This includes defining the influence graph and the process for constructing the local defense problems. We also discuss the scalability of our approach. In Section 4.5, we present an example illustrating some of the concepts used in our approach. In Section 4.6, we discuss our results and provide some concluding remarks.

4.1.2 Notation

The table below (presented for later reference) describes the notation used throughout the chapter.

$\mathscr{X} = \mathscr{X}^1 \times \mathscr{X}^2 \times \cdots \times \mathscr{X}^n$	state-space of the problem
$\mathscr{N} = \{1,2,\ldots,n\}$	set of state elements
$\mathscr{X}^i = \{x^{i,1}, x^{i,2}, \ldots, x^{i,n_x^i}\}$	state-space of element $i \in \mathscr{N}$
$\mathscr{T} = \{0,1,\ldots,T\}$	time horizon of length T
$\mathscr{W} = \{w^1, w^2, \ldots, w^{n_w}\}$	set of nature's events
$\mathscr{W}(x)$	set of nature's events admissible from state $x \in \mathscr{X}$
$\mathscr{U} = \mathscr{U}^1 \times \mathscr{U}^2 \times \cdots \times \mathscr{U}^n$	action-space
$\mathscr{U}^i = \{u^{i,1}, u^{i,2}, \ldots, u^{i,n_u^i}\}$	action-space of element $i \in \mathscr{N}$
$\mathscr{Y} = \mathscr{Y}^1 \times \mathscr{Y}^2 \times \cdots \times \mathscr{Y}^n$	set of event observations
$\mathscr{Y}^i = \{y^{i,1}, y^{i,2}, \ldots, y^{i,n_y^i}\}$	set of event observations for element i
$\mathscr{Z} = \mathscr{Z}^1 \times \mathscr{Z}^2 \times \cdots \times \mathscr{Z}^n$	set of action observations
$\mathscr{Z}^i = \{z^{i,1}, z^{i,2}, \ldots, z^{i,n_z^i}\}$	set of action observations for element i
$\mathscr{K} = \{1,2,\ldots,n_k\}$	set of local defense problems
$\mathscr{N}_k$	state indices of local defense problem k's internal state-space
$\mathscr{L}_k$	state indices of local defense problem k's local state-space
$\bar{\mathscr{N}_k}$	exogenous state indices for local defense problem k
$\mathscr{X}^{\mathscr{N}_k}$	internal state-space of local defense problem k
$\mathscr{X}^{\mathscr{L}_k}$	local state-space of local defense problem k
$\mathscr{X}^{\bar{\mathscr{N}_k}}$	exogenous state-space of local defense problem k
$\mathscr{U}^{\mathscr{N}_k}$	internal action space of local defense problem k
$\mathscr{Y}^{\mathscr{N}_k}$	internal event observation space of local defense problem k
$\mathscr{Z}^{\mathscr{N}_k}$	internal action observation space of local defense problem k
m_t^{kl}	message sent from local defense problem k to l at t
$m_t^{\bar{\mathscr{N}_k}}$	aggregate message of local defense problem k
$h_t = \{x_0, u_{0:t-1}, z_{0:t-1}, y_{0:t}\}$	realized history at time t
$h_t^k = \{x_0^{\mathscr{N}_k}, u_{0:t-1}^{\mathscr{N}_k}, z_{0:t-1}^{\mathscr{N}_k}, y_{0:t}^{\mathscr{N}_k}, m_{0:t}^{\bar{\mathscr{N}_k}}\}$	realized history for local defense problem k at time t
$\mathscr{H}_t$	space of histories for defender at time t
$\mathscr{H}_t^k$	space of histories for local defense problem k at time t
$\mathscr{R}$	space of information states
$\mathscr{B}^{\mathscr{L}_k}$	space of approximate information states for local defense problem k
ψ	information state update function
ϕ^k	approximate information state update of local defense problem k
$\mathscr{G}$	space of admissible defense policies

γ	element of $\mathscr{G}$, admissible defense policy
$\mathscr{G}^k$	space of admissible defense policies for local defense problem k
γ^k	element of $\mathscr{G}^k$, admissible defense policy for local defense problem k
Γ'^k	space of admissible approximate defense policies for local defense problem k
γ'^k	element of Γ'^k, admissible approximate defense policy for local defense problem k
$c(x,u)$	state-action cost
$c^i(x^i,u^i)$	state-action cost of element i
$x_{t:t+s}$	the sequence $x_t, x_{t+1}, \ldots, x_{t+s}$
$\mathscr{P}(\mathscr{A})$	the powerset of set $\mathscr{A}$
$v^{\mathscr{A}}$	the collection of elements v_i, $i \in \mathscr{A}$, from vector v
$v^{-i}, v^{-(i,j)}$	all elements of v excluding element i, resp. excluding elements i and j

We denote variables by upper-case letters and their realizations by their corresponding lower-case letter, *e.g.* x is a realization of variable X.

4.2 The Security Model

Consider a system consisting of n elements operating in discrete time. Let $\mathscr{N} := \{1,2,...,n\}$ denote the set of the system's elements. Consider a discrete, finite state space $\mathscr{X} := \mathscr{X}^1 \times \mathscr{X}^2 \cdots \times \mathscr{X}^n$, where $\mathscr{X}^i := \{x^{i,1}, x^{i,2}, \ldots, x^{i,n_x^i}\}$ represents the (finite) state-space of element $i \in \mathscr{N}$. Let T denote the time horizon over which we consider the system's operation; T may be finite or infinite. Define $\mathscr{T} = \{0,1,\ldots,T\}$. The state of the system at any given time t is given by

$$x_t = (x_t^1, x_t^2, \ldots, x_t^n) \in \mathscr{X}. \tag{4.1}$$

There are two agents: a controller (the defender) and nature (the adversary). The agents interact according to the following timing diagram, shown in Fig. 4.1.

The system state, x_t, evolves due to both the defender's decisions (control actions) and nature's events. For a given time-step (following the notation used in Fig. 4.1), nature first generates an event, $w_t(x_t)$; nature's set of feasible events at any time t depends on the system's state x_t, hence the notation $w_t(x_t)$. The set of all events due to nature is denoted by $\mathscr{W} := \{w^1, w^2, ..., w^{n_w}\}$. The set of events that are admissible from state x_t is denoted by $\mathscr{W}(x_t) \subseteq \mathscr{W}$. To simplify the notation in the

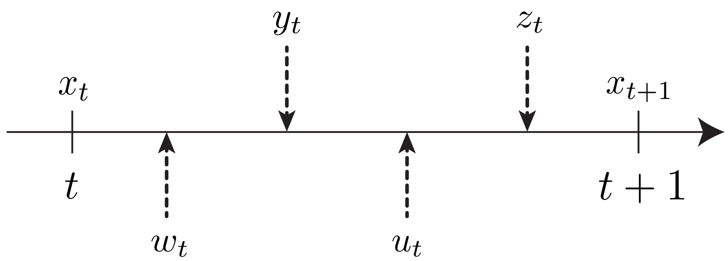


Figure 4.1: Timeline of events for a given time-step.

rest of the paper we use w_t instead of $w_t(x_t)$. The defender is not able to perfectly observe the event w_t, but instead receives an observation y_t, termed an *event observation*,generated according to the function $\theta : \mathscr{X} \times \mathscr{W} \rightarrow \mathscr{Y} = \mathscr{Y}^1 \times \mathscr{Y}^2 \times \cdots \times \mathscr{Y}^n$

$$\begin{aligned} y_t &= \theta(x_t, w_t) \\ &= \big(\theta^1(x_t^1, w_t), \ldots, \theta^n(x_t^n, w_t)\big) \end{aligned} \tag{4.2}$$

where $\theta^i : \mathscr{X}^i \times \mathscr{W} \rightarrow \mathscr{Y}^i$, $\mathscr{Y}^i = \{y^{i,1}, y^{i,2}, \ldots, y^{i,n_y^i}\}$, and n_y^i is the total number of possible observations for y_t^i. The defender then takes a defense action

$$u_t \in \mathscr{U} = \mathscr{U}^1 \times \mathscr{U}^2 \times \cdots \times \mathscr{U}^n \tag{4.3}$$

based on its current information, where it is assumed (for simplicity) that the action space decomposes over the elements of the state-space. Each space $\mathscr{U}^i$ consists of a finite set of defense alternatives $\mathscr{U}^i = \{u^{i,1}, u^{i,2}, \ldots, u^{i,n_u^i}\}$. We assume that, for a given element i, each action $u^{i,j} \in \mathscr{U}^i$, $j \in \{1, 2, \ldots, n_u^i\}$, only has an effect on the state x^i of element i. The defender incurs a state-dependent cost for each defense action u_t, denoted by

$$c(x_t, u_t) = \sum_{i \in \mathscr{N}} c^i(x_t^i, u_t^i), \tag{4.4}$$

We assume that $|c(x,u)| \leq c^{\max}$ for all $x \in \mathscr{X}$ and $u \in \mathscr{U}$. This cost is incurred immediately after the defense action is selected. The defender then receives an observation z_t, termed an *action observation*, following the defense action, that is generated by the function $\zeta : \mathscr{X} \times \mathscr{W} \times \mathscr{U} \rightarrow \mathscr{Z} = \mathscr{Z}^1 \times \mathscr{Z}^2 \times \cdots \times \mathscr{Z}^n$ as

$$\begin{aligned} z_t &= \zeta(x_t, w_t, u_t) \\ &= \big(\zeta^1(x_t^1, w_t, u_t^1), \ldots, \zeta^n(x_t^n, w_t, u_t^n)\big) \end{aligned} \tag{4.5}$$

where $\zeta^i : \mathscr{X}^i \times \mathscr{W} \times \mathscr{U}^i \to \mathscr{Z}^i$, $\mathscr{Z}^i = \{z^{i,1}, z^{i,2}, \ldots, z^{i,n_z^i}\}$, and n_z^i is the total number of possible observations for z_t^i. The defense action, $u_t = (u_t^1, \ldots, u_t^n)$, causes the system to transition to the system state x_{t+1} according to the state update function $\pi : \mathscr{X} \times \mathscr{W} \times \mathscr{U} \to \mathscr{X}$, that is,

$$\begin{aligned} x_{t+1} &= \pi(x_t, w_t, u_t) \\ &= \big(\pi^1(x_t, w_t, u_t^1), \ldots, \pi^n(x_t, w_t, u_t^n)\big). \end{aligned} \tag{4.6}$$

where each $\pi^i : \mathscr{X} \times \mathscr{W} \times \mathscr{U}^i \to \mathscr{X}^i$ is the update equation state element i. Note that at any time t, each π^i depends on the global system state x_t, not only on x_t^i, because, as we pointed out above, the set of nature's feasible events at t depends on x_t.

4.3 The Defense Problem

The optimal defense action at any given time-step is dictated by an optimal defense policy. The defense policy at time t, denoted by γ_t, is a function of the defender's information available at time t. This information, termed the *history* and denoted by h_t, consists of the initial state x_0, all previous control actions, $u_0, \ldots, u_{t-1}$ (denoted compactly by $u_{0:t-1}$), and all observations $y_{0:t}$, and $z_{0:t-1}$. Formally, $h_t = \{x_0, u_{0:t-1}, z_{0:t-1}, y_{0:t}\}$, where the initial state x_0 is known by the defender. A defense policy, $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_T)$, maps the available information at any time t to a defense action u_t, that is, $\gamma_t : \mathscr{H}_t \to \mathscr{U}$, where $\mathscr{H}_t$ is the information space of the controller at time t (the space of histories up to time t). The space of admissible defense policies is given by $\mathscr{G} = \{\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_T) \,|\, \gamma_t : \mathscr{H}_t \to \mathscr{U} \text{ for all } t \in \mathscr{T}\}$.

An optimal defense policy is a policy γ that solves the following partially observable minmax control problem (P).

$$\min_{\gamma \in \mathscr{G}} \max_{\{X_{\mathscr{T}} \in \mathscr{X}_{\mathscr{T}}^{\gamma}\}} \left\{ \sum_{t \in \mathscr{T}} \beta^t c(X_t, U_t) \,\middle|\, X_0 = x_0 \right\} \tag{P}$$

$$\text{subject to } \; X_{t+1} = \pi(X_t, W_t, U_t) \tag{P-i}$$

$$Y_t = \theta(X_t, W_t) \tag{P-ii}$$

$$Z_t = \zeta(X_t, W_t, U_t) \tag{P-iii}$$

$$U_t = \gamma_t(H_t) \tag{P-iv}$$

$$H_t = \{x_0, U_{0:t-1}, Z_{0:t-1}, Y_{0:t}\} \tag{P-v}$$

The set $\mathscr{X}_{\mathscr{T}}^{\gamma}$ denotes the space of all sequences of system states (trajectory) generated by a given defense policy γ. The maximization is taken over all families of state trajectories generated by the defense policy γ, denoted by $\{X_{\mathscr{T}} = \{X_1, X_2, \ldots, X_T\} \in \mathscr{X}_{\mathscr{T}}^{\gamma}\}$. We consider all such families of trajectories (each one associated with a defense policy) and choose the policy that minimizes the highest-cost (worst-case) trajectory among all families.

The remainder of this section is devoted to determining more compact descriptions of the defender's information. Such descriptions may either permit the computation of an optimal defense policy or provide guidelines/insights for computationally tractable approximations of Problem (P).

4.3.1 Information State

In order to prescribe an optimal defense action at time t for Problem (P), we need to determine an *information state* sufficient for performance evaluation. One such information state is the history h_t. Unfortunately, due to the unbounded growth of the domain of h_t (see Fig. 4.2), the computation of a defense policy based on h_t is intractable for infinite-time horizon problems (and large finite horizons). This motivates the search for a more compact (albeit still sufficient) summary of the current information.[2]

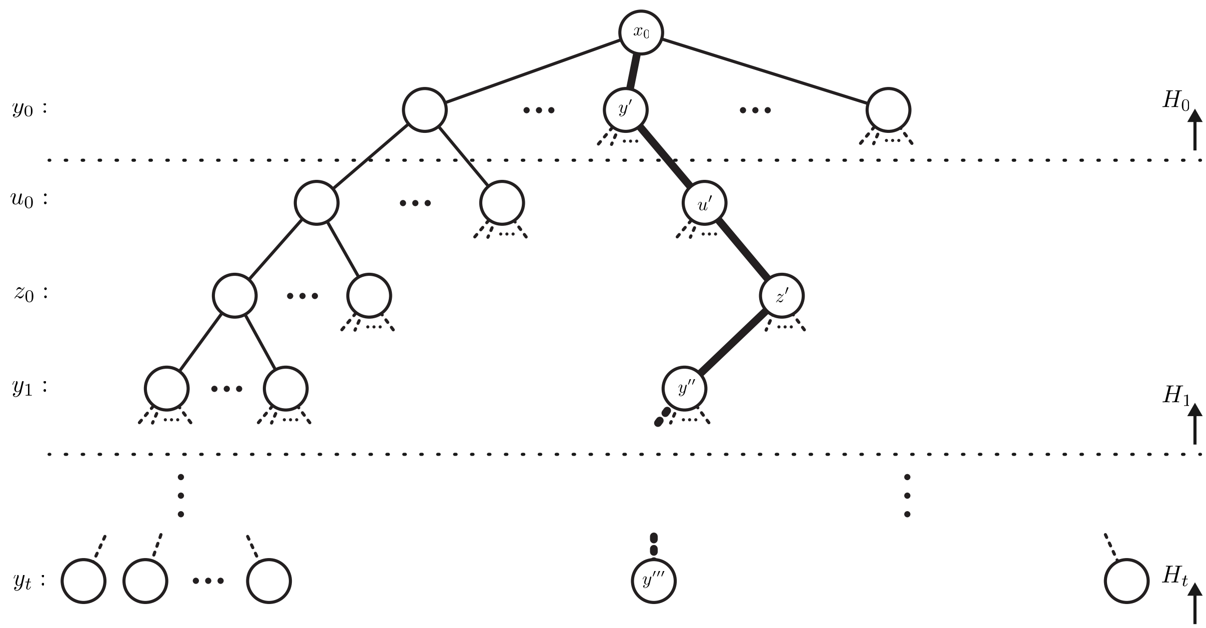


Figure 4.2: The information, H_t, consisting of the initial state x_0, observations $Y_{0:t}$, $Z_{0:t-1}$, and previous defense actions $U_{0:t-1}$ grows rapidly as a function of t, as can be seen by the tree structure above. The number of information realizations at time t is equal to the number of leaf nodes in the tree at depth t. A *realized information trajectory*, $h_t = (x_0, y', u', z', y'', \ldots, y''')$, is a path from the root of the tree, x_0, to a leaf, in this case y''', as shown by the bolded line.

[2]The need for finding a compact information state is even more critical in modern dynamic security environments where the rate of events is high, causing h_t to grow rapidly in size.

An alternate information state for Problem (P) can be defined. This alternate information state offers insight into the source of computational difficulty associated with the defender's problem and forms the basis for later approximations that bring the problem into the realm of computational tractability.

We now define this information state, denoted by R_t, in the context of the defense problem and describe its update equations. Consider a realization $h_t := \{x_0, u_{0:t-1}, z_{0:t-1}, y_{0:t}\}$ of the defender's information at time t. Denote by $\{\mathscr{J}_t^1, \ldots, \mathscr{J}_t^{\hat{l}(h_t)}\}$ to be the set of distinct event-action trajectories from 0 to t that are compatible with x_0 and h_t;[3] $\mathscr{J}_t^i = \{x_0, u_{0:t-1}, {}^i w_{0:t}\}$, $i = 1, 2, \ldots, \hat{l}(h_t)$. Each trajectory $\mathscr{J}_t^i$ leads to a system state ${}^i\hat{x}_t := ({}^i\hat{x}_t^1, \ldots, {}^i\hat{x}_t^{i,n}), i = 1, 2, \ldots, l(h_t)$. Denote by ${}^i\hat{\kappa}_t$ the cost of reaching ${}^i\hat{x}_t$; ${}^i\hat{\kappa}_t := \sum_{\tau=0}^{t-1} \beta^\tau c({}^i\hat{x}_\tau, u_\tau)$, where ${}^i\hat{x}_\tau$ is the system state at τ due to $\mathscr{J}_t^i$. Define $\hat{R}_t(h_t) := \{({}^1x_t, {}^1w_t, {}^1\hat{\kappa}_t), \ldots, ({}^{\hat{l}(h_t)}x_t, {}^{\hat{l}(h_t)}w_t, {}^{\hat{l}(h_t)}\hat{\kappa}_t)\}$. Apply the following *reduction process* to $\hat{R}(h_t)$. If $\hat{R}(h_t)$ contains components $({}^{j_1}x_t, {}^{j_1}w_t, {}^{j_1}\kappa_t), \ldots, ({}^{j_q}x_t, {}^{j_q}w_t, {}^{j_q}\kappa_t)$ such that $({}^{j_1}x_t, {}^{j_1}w_t) = \ldots = ({}^{j_q}x_t, {}^{j_q}w_t)$ and ${}^{j_p}\hat{\kappa}_t = \max_{j \in \{j_0, j_1, \ldots, j_q\}} \{{}^j\kappa_t\}$, then omit $({}^{j_i}x_t, {}^{j_i}w_t, {}^{j_i}\kappa_t)$, $j_i \neq j_p$, from $\hat{R}_t(h_t)$. This reduction process results in

$$R_t(h_t) = \{({}^1x_t, {}^1w_t, {}^1\kappa_t), \ldots, ({}^{l(h_t)}x_t, {}^{l(h_t)}w_t, {}^{l(h_t)}\kappa_t)\}$$

where $({}^1x_t, {}^1w_t), \ldots, ({}^{l(h_t)}x_t, {}^{l(h_t)}w_t)$ are distinct; $R_t(h_t)$ is an alternative information state at t along h_t for Problem (P). From the construction of $R_t(h_t)$, we conclude that for all history realizations h_t and for all t, an information state R_t has the form $\{({}^1x_t, {}^1w_t, {}^1\kappa_t), ..., ({}^ax_t, {}^aw_t, {}^a\kappa_t)\}$, where $\{({}^1x_t, {}^1w_t), ..., ({}^ax_t, {}^aw_t)\} \in \mathscr{P}(\mathscr{X} \times \mathscr{W}^n) = \mathscr{P}(\mathscr{X}^1 \times \mathscr{X}^2 \times ... \times \mathscr{X}^n \times \mathscr{W}^n)$, ${}^i\kappa_t \in [0, \frac{c^{\max}}{1-\beta}]$, for all i, for all t, and $\mathscr{P}(\cdot)$ denotes the power set. We denote by $\mathscr{R}$ the space of information states at any time t.

We now describe how to obtain R_{t+1} from R_t and the new information $H_{t:t+1}$ that becomes available to the defender at time $t+1$. Let h_t be a realization of H_t and $h_{t:t+1} = h_{t+1} \backslash h_t = \{u_t, z_t, y_{t+1}\}$ be a realization of $H_{t:t+1}$. Denote by $\{\mathscr{J}_{t:t+1}^1, ..., \mathscr{J}_{t:t+1}^{p(h_{t:t+1})}\}$ the set of distinct action-event trajectories from t to $t+1$ that are compatible with $R(h_t)$ and $h_{t:t+1}$; $\mathscr{J}_{t:t+1}^i := \{u_t, {}^iw_{t+1}\}$, $i = 1, 2, ..., p(h_{t:t+1})$. Using $R(h_t)$, $\{\mathscr{J}_{t:t+1}^1, \cdots \mathscr{J}_{t:t+1}^{p(h_{t:t+1})}\}$ and the system dynamics, Eq. (4.6), we can construct $\hat{R}(h_{t+1}) := \{({}^r\hat{x}_{t+1}, {}^r\hat{w}_{t+1}, {}^r\hat{\kappa}_{t+1}), r = 1, 2, \ldots, a\}$[4] where ${}^r\hat{x}_{t+1} = \pi(({}^rx_t, {}^rw_t, u_t)$, $(u_t, {}^rw_{t+1}) \in \{\mathscr{J}_{t:t+1}^1, \cdots \mathscr{J}_{t:t+1}^{p(h_{t:t+1})}\}$, ${}^rw_{t+1} \in \mathscr{W}({}^r\hat{x}_{t+1})$, ${}^r\hat{\kappa}_{t+1} = {}^r\kappa_t + \beta^t c({}^rx_t, u_t)$, and ${}^r\kappa_t$ is the cost associated with $({}^rx_t, {}^rw_t)$, where $({}^rx_t, {}^rw_t, {}^r\kappa_t) \in R(h_t)$. Apply the above-described reduction

[3]Compatibility implies that the following requirements are satisfied; each event-action trajectory $\{u_{0:t-1}, {}^iw_{0:t}\}$ is consistent with x_0, the observations $y_{0:t}$ and $z_{0:t-1}$ (through θ and ζ, respectively), the system dynamics described by Eq. (4.6), and for every $\tau \leq t$, $w_\tau \in \mathscr{W}(x_\tau)$, where x_τ is the system state at τ reached via x_0, $u_{0:\tau-1}$, $w_{0:\tau-1}$.

[4]The number of components a of $\hat{R}(h_{t+1})$ depends on $R(h_t)$, $h_{t:t+1}$, the system dynamics, and the sets $\mathscr{W}(x)$, $x \in \mathscr{X}$.

process to $\hat{R}_{t+1}(h_{t+1})$ to obtain the alternative information state

$$R_{t+1}(h_{t+1}) = \{({}^1x_{t+1}, {}^1w_{t+1}, {}^1\kappa_{t+1}), \ldots, ({}^{l(h_{t+1})}x_{t+1}, {}^{l(h_{t+1})}w_{t+1}, {}^{l(h_{t+1})}\kappa_{t+1})\}$$

for Problem (P) at time $t+1$ along h_{t+1}. The recursive update process described above can be summarized by an update equation

$$R_{t+1}(h_{t+1}) = \psi(R_t(h_t), h_{t:t+1}) = \psi(R_t(h_t), u_t, z_t, y_{t+1}). \tag{4.7}$$

The information state described above summarizes the information h_t available to the defender at time t by including all system states at t that are compatible with h_t, the maximum cost that is incurred in order to reach each of these states, and the events in nature that follow each of the possible states at t; these events must be feasible conditioned on the state ${}^rx_t$, that is, ${}^rw_t$ must be in the set $W({}^rx_t)$.

We are now in a position to write the dynamic program for Problem (P).

4.3.2 Sequential Decomposition and Dynamic Programming

We discuss a sequential decomposition for Problem (P) using dynamic programming. To specify the dynamic program for the finite horizon problem, denote by r_t the information state at time $t \le T$. Define by $V_t(r_t)$, $r_t \in R_t$, the minmax value of Problem (P) from time t on when the information state at t is r_t. Then, when $r_T = (({}^1x, {}^1w, {}^1\kappa), ({}^2x, {}^2w, {}^2\kappa), \ldots, ({}^{l(h_T)}x, {}^{l(h_T)}w, {}^{l(h_T)}\kappa))$,

$$V_T(r_T) = \max_{j \in \{1,2,\ldots,l(h_T)\}} {}^j\kappa \tag{4.8}$$

For $t = 0, 1, \ldots, T-1$,

$$\begin{aligned} V_t(r_t) &= \min_{u_t \in \mathscr{U}} \left[\max_{(x_t, w_t, \kappa_t) \in r_t} V_{t+1}(r_{t+1}) \right] \\ &= \min_{u_t \in \mathscr{U}} \Bigg[\max_{(x_t, w_t, \kappa_t) \in r_t} \Big[\max_{w_{t+1} \in \mathscr{W}(\pi(x_t, w_t, u_t))} \\ &\qquad V_{t+1}\Big(\psi_t\big(r_t, u_t, \zeta(x_t, w_t, u_t), \theta(\pi(x_t, w_t, u_t), w_{t+1})\big)\Big)\Big]\Bigg]. \end{aligned} \tag{4.9}$$

Eqs. (4.8) and (4.9) define the dynamic program for the finite (T) horizon Problem (P).

To specify the dynamic program for the infinite horizon problem, we let $r \in \mathscr{R}$ denote the current information state and $V(r)$ denote the minmax value of the infinite horizon Problem (P).

Then,

$$V(r) = \min_{u \in \mathscr{U}} \left[\max_{(x,w,\kappa) \in r} \left[\max_{w' \in \mathscr{W}(\pi(x,w,u))} V\Big(\psi\big(r, u, \zeta(x,w,u), \theta(\pi(x,w,u), w')\big)\Big)\right]\right] \tag{4.10}$$

Because of the high dimensionality, the solution of the finite and infinite time-horizon dynamic programs is computationally intractable. For this reason, in the next section, we provide a scalable approach for the solution of Problem (P).

4.4 Approximation to the Defense Problem

Even though the alternate information state described in Section 4.3.1 above leads to an intractable problem even for small systems, it forms the basis for a scalable approach to the solution of Problem (P). The approach consists of two key steps: (i) Using the concept of an *influence graph*, we analyze the functional dependencies between state elements and split elements with weak dependencies, decomposing the original problem (P) into a number of local defense problems (P_k); (ii) We further approximate the solution of each local defense problem in order to permit computation of (suboptimal) local defense policies. We discuss the computational complexity of each of the local defense problems and comment on how to use the features of our approach so as to end up with problems that are compatible with the defender's computational capabilities.

4.4.1 Local Defense Problems

4.4.1.1 Preliminaries

In order to define the local defense problems, denoted by (P_k), we first introduce some necessary notation and describe, at a high level, how the local defense problems interact with one another. Then, in the remaining subsections, we describe in detail how the local defense problems are formed.

Consider a collection of n_k local defense problems, denoted by the set $\mathscr{K} = \{1, 2, \dots, n_k\}$. Each local defense problem $k \in \mathscr{K}$ has an associated set of states termed the *internal state-space* of problem k, denoted by $\mathscr{X}^{\mathscr{N}_k} \subseteq \mathscr{X}$, where $\mathscr{N}_k \subseteq \mathscr{N}$ is the set of internal state indices for problem k. By construction (described later), $\mathscr{X}^{\mathscr{N}_k}$, $k \in \mathscr{K}$, form a partition of the original state-space $\mathscr{X}$, that is $\mathscr{X}^{\mathscr{N}_i} \cap \mathscr{X}^{\mathscr{N}_j} = \emptyset$ for $i \neq j$, and $\cup_{k \in \mathscr{K}} \mathscr{X}^{\mathscr{N}_k} = \mathscr{X}$. Under this partition, the action and observation spaces for each local defense problem $k \in \mathscr{K}$ are denoted by $\mathscr{U}^{\mathscr{N}_k}$ and $\mathscr{Y}^{\mathscr{N}_k}$, $\mathscr{Z}^{\mathscr{N}_k}$, respectively.

Each local defense problem (P_k) is associated with a local defense policy γ^k. In computing the local defense policies, we assume that the local defense problems can exchange information over time via messages. We denote by m_t^{kl} the message local defense problem k receives from local defense problem l at time t. The message exchange process occurs immediately before each local defense action is taken, as shown in the timing diagram of Fig. 4.3.

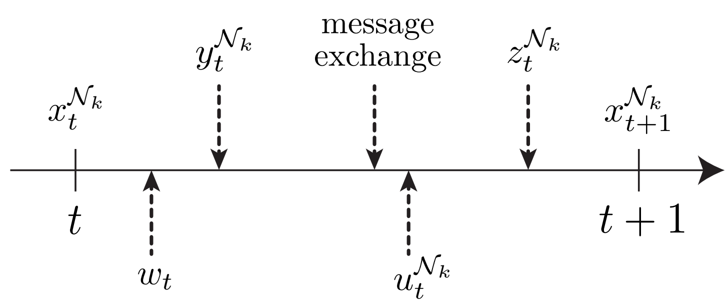


Figure 4.3: Timeline of events for the local defense problem (P_k) for a given time-step.

The local defense action, $u_t^{\mathscr{N}_k} \in \mathscr{U}^{\mathscr{N}_k}$, causes the internal states of local defense problem k, denoted by $x_t^{\mathscr{N}_k} = \{x_t^j \mid j \in \mathscr{N}_k\}$, to transition to $x_{t+1}^{\mathscr{N}_k}$ according to the state update function $\pi^{\mathscr{N}_k} : \mathscr{X} \times \mathscr{W} \times \mathscr{U}^{\mathscr{N}_k} \to \mathscr{X}^{\mathscr{N}_k}$. The function $\pi^{\mathscr{N}_k}$ is simply defined as the collection of functions π^j, $j \in \mathscr{N}_k$, as described in Eq. (4.6). Note that the dynamics of the internal states of each local defense problem (P_k) depend, in general, on the state of the overall system. This means that, without exploiting any additional structure of the problem, the state space of each local defense problem is just as large as that of the original problem (P). To address this, we define an influence graph (in Section 4.4.1.2) in order to quantify the functional dependencies among state elements. Using the structure of the influence graph, each local defense problem can restrict attention to the state elements that directly influence the evolution of its internal states. The remainder of Section 4.4 is devoted to describing how the local defense problems are formed.

4.4.1.2 Functional Dependencies & the Notion of an Influence Graph

To form the local defense problems, we first analyze the functional dependencies among the state elements $i \in \mathscr{N}$. To this end, we provide the following definition.

Definition 4.4.1 (Functional Dependency)**.** *State element i is said to have a* functional dependency *on state element j, if there exists an action $u^i \in \mathscr{U}^i$, an event $w \in \mathscr{W}$, and two states $x = (x^1, \ldots, x^{j-1}, x^j, x^{j+1}, \ldots, x^n)$, $\hat{x} = (x^1, \ldots, x^{j-1}, \hat{x}^j, x^{j+1}, \ldots, x^n)$ differing only in element j, $x^j \neq \hat{x}^j$, such that*

$$\pi^i(x, w, u^i) \neq \pi^i(\hat{x}, w, u^i),$$

where π^i is the state update function of element i, given by Eq. (4.6).

In other words, state element i is said to be functionally dependent on state element j if a change in the state element j influences the update for state element i for some action $u^i \in \mathscr{U}^i$ and some event $w \in \mathscr{W}$. The relationships expressed by Definition 4.4.1 can be summarized by a graph, termed the *influence graph*, defined below.

Definition 4.4.2 (Influence Graph)**.** *The* influence graph*, $\mathscr{G} = (\mathscr{V}, \mathscr{E}, \xi)$, is a weighted directed graph, that consists of nodes $\mathscr{V} = \mathscr{N}$, edge set $\mathscr{E}$, and edge-weights ξ. The edge-set $\mathscr{E}$ contains edge $e = (i, j)$ if and only if state element j is functionally dependent on state element i, as described by Definition 4.4.1. Each edge $e = (i, j) \in \mathscr{E}$ in the influence graph is assigned a weight $\xi_e \in (0, 1]$, computed as*

$$\xi_e = \frac{1}{d} \sum_{\substack{x^{-(i,j)} \in \mathscr{X}^{-(i,j)} \\ x^i \in \mathscr{X}^i \\ x^j \in \mathscr{X}^j \\ w \in \mathscr{W} \\ u^i \in \mathscr{U}^i \\ \hat{x}^j \in \mathscr{X}^j \setminus \{x^j\}}} \mathbf{1}\left(\{\pi^i\left(x, w, u^i\right) \neq \pi^i\left(\hat{x}, w, u^i\right)\}\right) \tag{4.11}$$

where $\mathbf{1}(A)$ is the indicator function of event A and d is the normalization term; d is the number of all possible terms in the summation of Eq. (4.11) and therefore, its value is equal to $d = (\prod_{l \neq i,j} n_x^l) n_x^i n_x^j n_w n_u^i (n_x^j - 1)$.

The influence graph summarizes the dependencies among all state elements for all possible defense actions and events. A directed edge $e = (i, j) \in \mathscr{E}$ exists from element (node) i to element j if there exists an assignment of values to states, events, and actions such that state element j functionally depends on state element i. The weight of edge $(i, j) \in \mathscr{E}$ quantifies the strength of the functional dependency of element j on element i.

4.4.1.3 Formulating the Local Defense Problems

We form local defense problems by partitioning the influence graph into *clusters* of nodes (state elements). The clustering algorithm should place state elements with strong dependencies into a single cluster. Since the weights of the edges in the influence graph quantify the strength of the functional dependencies between state elements, application of the (normalized) min-cut algorithm Johnson et al. (1993) splits elements with weak functional dependencies, achieving the desired goal. A single application of the (normalized) min-cut algorithm decomposes a graph into two separate graphs using a minimum weight cut. Continued application of the min-cut algorithm results in increasingly finer partitioning of nodes, resulting in clusters with fewer nodes and eventually leading to a partitioning of the state-space that permits computation of local defense policies (this procedure is described in detail in the example of Section 4.5). For each cluster of elements, we formulate a local defense problem (P_k) and aim to solve for the corresponding local defense policy.

In order to formulate each local defense problem (P_k), $k \in \mathscr{K}$, we must first describe its associated state-space and dynamics. Using the influence graph, one can analyze the functional dependencies between state elements. Specifically, given a set of internal state indices $\mathscr{N}_k$, one can determine the state elements outside the set of internal states, $i \in \mathscr{N} \setminus \mathscr{N}_k$, that influence the evolution of states within $\mathscr{N}_k$ (as described by Definition 4.4.1). Formally, this set of elements is defined as

$$\bar{\mathscr{N}_k} = \{i \in \mathscr{N} \setminus \mathscr{N}_k \mid (i,j) \in \mathscr{E}, j \in \mathscr{N}_k\}$$

where $\mathscr{E}$ is the influence graph's set of edges. From this, one can define the set of *local* state element indices as the state elements in $\mathscr{N}_k$ combined with the state elements that directly influence the evolution of any element in $\mathscr{N}_k$, that is,

$$\mathscr{L}_k = \mathscr{N}_k \cup \bar{\mathscr{N}_k}.$$

The state-space corresponding to indices $\mathscr{L}_k$ is written as $\mathscr{X}^{\mathscr{L}_k} \subseteq \mathscr{X}$ and is termed the *local state-space* of problem k. Since, by construction, there are no edges from any elements in $\mathscr{N} \setminus \mathscr{L}_k$ to elements in $\mathscr{N}_k$, the update of the internal states $x^{\mathscr{N}_k}$ is independent of the states in $\mathscr{X}^{\mathscr{N} \setminus \mathscr{L}_k}$. Therefore, with some abuse of notation, the state update function for the internal states can equivalently be written using the function $\pi^{\mathscr{N}_k} : \mathscr{X}^{\mathscr{L}_k} \times \mathscr{W} \times \mathscr{U}^{\mathscr{N}_k} \to \mathscr{X}^{\mathscr{N}_k}$ as

$$x_{t+1}^{\mathscr{N}_k} = \pi^{\mathscr{N}_k}(x_t^{\mathscr{L}_k}, w_t, u_t^{\mathscr{N}_k}). \tag{4.12}$$

It is important to note that the local state-space $\mathscr{X}^{\mathscr{L}_k}$ of each local defense problem $k \in \mathscr{K}$ is an approximation to the state-space that would be necessary for computing an optimal local defense policy. Although the states $x^{\mathscr{N}_k}$ can be updated precisely using the local states $x^{\mathscr{L}_k}$, as evidenced by Eq. (4.12), computation of an optimal local defense policy requires knowledge of how $x^{\mathscr{L}_k}$ evolves over time. Due to the computational limitations of the defender, one is unable to take into account the state dynamics associated with the state elements outside its internal state-space, $\mathscr{X}^{\mathscr{N}_k}$, for the purpose of computing a local defense policy. As a result, we approximate the state dynamics by introducing a message-exchange procedure in which *neighboring* defense problems communicate relevant elements of their internal states.

The messages that local defense problems exchange consist of the possible states that are consistent with each local defense problem's history (more on this in Section 4.4.2). Before formally defining the messages, define the set

$$\bar{\mathcal{N}_{kl}} = \bar{\mathcal{N}_k} \cap \mathcal{N}_l.$$

The set $\bar{\mathcal{N}_{kl}}$ represents the set of state indices within $\mathcal{N}_l$ that can influence the evolution of a state element in $\mathcal{N}_k$. Note that $\bar{\mathcal{N}_{kl}}$ is only non-empty if there is an edge $(i,j) \in \mathcal{E}$ in the influence graph such that $i \in \mathcal{N}_l$ and $j \in \mathcal{N}_k$. Also, note that $\{\bar{\mathcal{N}_{kl}}, l \in \mathcal{K}\}$ forms a partition of $\bar{\mathcal{N}_k}$. The message that local defense problem k receives from local defense problem l at time t, m_t^{kl}, lives within the set of all possible states of $x^{\bar{\mathcal{N}_{kl}}}$, that is, $m_t^{kl} \in \mathcal{P}(\mathcal{X}^{\bar{\mathcal{N}_{kl}}})$. Local defense problem l constructs this message as the set of possible states that are consistent with its local information (built in part using its imperfect internal observations).

In this sense, local defense problem k receives a summary of local defense problem l's local information that is relevant for taking an internal defense action $u^{\mathcal{N}_k} = \{u^i \mid i \in \mathcal{N}_k\} \in \mathcal{U}^{\mathcal{N}_k} = \prod_{i \in \mathcal{N}_k} \mathcal{U}^i$, permitting local defense problem k to compute a (suboptimal) local defense policy γ^k. The complete set of messages that local defense problem k receives from neighboring local defense problems are combined to form an *aggregate message*, denoted by

$$m_t^{\bar{\mathcal{N}_k}} \in \mathcal{P}\left(\prod_{l \in \mathcal{K}} \mathcal{X}^{\bar{\mathcal{N}_{kl}}}\right) = \mathcal{P}(\mathcal{X}^{\bar{\mathcal{N}_k}}).$$

The aggregate message allows local defense problem k to update its exogenous state elements in a way that is consistent with the exchanged information, that is, $x_{t+1}^{\bar{\mathcal{N}_k}} \in m_{t+1}^{\bar{\mathcal{N}_k}}$. Combined with the state update of its internal states, given by Eq. (4.12), the defender is able to (approximately) model the evolution of its local states $x^{\mathcal{L}_k}$.

For each local defense problem $k \in \mathcal{K}$, an optimal local defense action at any given time-step is dictated by an optimal local defense policy. The local defense policy at time t, denoted by γ_t^k, prescribes an action based on its available information at time t. This information, given by $h_t^k = \{x_0^{\mathcal{N}_k}, u_{0:t-1}^{\mathcal{N}_k}, z_{0:t-1}^{\mathcal{N}_k}, y_{0:t}^{\mathcal{N}_k}, m_{0:t}^{\bar{\mathcal{N}_k}}\}$, consists of the initial value of the internal states $x_0^{\mathcal{N}_k}$, all internal defense actions $u_{0:t-1}^{\mathcal{N}_k}$ up to and including $t-1$, all internal action observations $z_{0:t-1}^{\mathcal{N}_k} = \{z_{0:t-1}^i \mid i \in \mathcal{N}_k\}$ up to and including $t-1$, all internal event observations $y_{0:t}^{\mathcal{N}_k} = \{y_{0:t}^i \mid i \in \mathcal{N}_k\}$ up to and including t, and all received messages up to and including t, summarized by the aggregate messages $m_{0:t}^{\bar{\mathcal{N}_k}}$. A local defense policy, $\gamma^k = (\gamma_1^k, \gamma_2^k, \ldots, \gamma_T^k)$, maps the available information at any time t, h_t^k, to a defense action $u_t^{\mathcal{N}_k}$, that is, $\gamma_t^k : \mathcal{H}_t^k \to \mathcal{U}^{\mathcal{N}_k}$, where $\mathcal{H}_t^k$ is the information space of the local defense problem at time t. The space of admissible local defense policies for local defense problem k is given by $\mathcal{G}^k = \{\gamma^k = (\gamma_1^k, \gamma_2^k, \ldots, \gamma_T^k) \mid \gamma_t^k : \mathcal{H}_t^k \to \mathcal{U}^{\mathcal{N}_k} \text{ for all } t \in \mathcal{T}\}$. The optimal local defense policy for local defense problem k, $k \in \mathcal{K}$, is a policy γ^k that solves the following partially observable minimax control problem (P_k).

$$\min_{\gamma^k \in \mathscr{G}^k} \max_{\left\{X_{\mathscr{T}}^{\mathscr{L}_k} \in \mathscr{X}_{\mathscr{T}}^{\mathscr{L}_k,\gamma^k}\right\}} \left\{ \sum_{t \in \mathscr{T}} \beta^t c^{\mathscr{N}_k}(X_t^{\mathscr{N}_k}, U_t^{\mathscr{N}_k}) \,\middle|\, X_0^{\mathscr{L}_k} = x_0^{\mathscr{L}_k} \right\} \tag{P_k}$$

$$\text{subject to } \; X_{t+1}^{\mathscr{N}_k} = \pi^{\mathscr{N}_k}(X_t^{\mathscr{L}_k}, W_t, U_t^{\mathscr{N}_k}) \tag{P_k-i}$$

$$X_{t+1}^{\bar{\mathscr{N}}_k} \in M_{t+1}^{\bar{\mathscr{N}}_k} \tag{P_k-ii}$$

$$Y_t^{\mathscr{N}_k} = \theta^{\mathscr{N}_k}(X_t^{\mathscr{N}_k}, W_t) \tag{P_k-iii}$$

$$Z_t^{\mathscr{N}_k} = \zeta^{\mathscr{N}_k}(X_t^{\mathscr{N}_k}, W_t, U_t^{\mathscr{N}_k}) \tag{P_k-iv}$$

$$U_t^{\mathscr{N}_k} = \gamma^k(H_t^k) \tag{P_k-v}$$

where the local states are defined as collection of internal and exogenous states, $X_t^{\mathscr{L}_k} = (X_t^{\mathscr{N}_k}, X_t^{\bar{\mathscr{N}}_k})$. The functions $\theta^{\mathscr{N}_k}$, $\zeta^{\mathscr{N}_k}$ are defined as the collection of event and action observation functions θ^i, ζ^i, $i \in \mathscr{N}_k$, as defined in Eqs. (4.2) and (4.5), respectively. The function $c^{\mathscr{N}_k}(x^{\mathscr{N}_k}, u^{\mathscr{N}_k})$ represents the state-action cost of local defense problem k and is defined as the sum of the internal state-action cost functions, that is, $c^{\mathscr{N}_k}(x^{\mathscr{N}_k}, u^{\mathscr{N}_k}) = \sum_{j \in \mathscr{N}_k} c^j(x^j, u^j)$. The set $\mathscr{X}_{\mathscr{T}}^{\mathscr{L}_k,\gamma^k}$ denotes the space of all sequences of local system states under local defense policy γ^k. As described earlier for Problem (P), the optimal policy is the policy that minimizes the worst-case cost over all state trajectories.

Each local defense problem (P_k), defined above, is of the same form as Problem (P). Consequently, the information state for the internal state-space $\mathscr{X}^{\mathscr{N}_k}$ is of the same form as the one described in Section 4.3.1. Unfortunately, such an information state precludes computation of an optimal local defense policy. As a result, we approximate the information state of each local defense problem (P_k) so as to end up with a modified problem that is computationally tractable. This is the topic of the following subsection.

4.4.2 Approximating the Local Defense Problems

We use the information state described in Section 4.4.1 to form an approximate information state for each local defense problem k. For a given realization of the local history $h_t^k \in \mathscr{H}_t^k$, we consider only the *set of states* that are compatible with h_t^k.[5] That is, each local defense problem constructs the set of all possible local states, $x^{\mathscr{L}_k} = (x^{\mathscr{N}_k}, x^{\bar{\mathscr{N}}_k})$, consistent with the history of internal defense actions and observations and messages that it has received from neighboring local defense problems. We denote this approximate information state at time t by $b_t^{\mathscr{L}_k} \in \mathscr{B}^{\mathscr{L}_k} = \mathscr{P}(\mathscr{X}^{\mathscr{L}_k})$, where $\mathscr{B}^{\mathscr{L}_k}$ is the space of approximate information states for local defense problem k. Using the new information from time t to $t+1$, given by $h_{t:t+1}^k = \{u_t^{\mathscr{N}_k}, z_t^{\mathscr{N}_k}, y_{t+1}^{\mathscr{N}_k}, m_{t+1}^{\bar{\mathscr{N}}_k}\}$, local defense problem k can update its approximate

[5] We omit the set of nature's events and the maximum cost associated with each compatible state.

information state as follows. We note that $b_t^{\mathscr{L}_k}$ has the form $b_t^{\mathscr{L}_k} = \{({}^1x_t^{\mathscr{N}_k}, {}^1x_t^{\bar{\mathscr{N}_k}}), \dots, ({}^lx_t^{\mathscr{N}_k}, {}^lx_t^{\bar{\mathscr{N}_k}})\}$. We use $({}^1x_t^{\mathscr{N}_k}, \dots, {}^lx_t^{\mathscr{N}_k})$, $u_t^{\mathscr{N}_k}$, $z_t^{\mathscr{N}_k}$, $y_{t+1}^{\mathscr{N}_k}$ to determine $({}^1x_{t+1}^{\mathscr{N}_k}, \dots, {}^rx_{t+1}^{\mathscr{N}_k})$ according to the update process ψ described in Section 4.3.1. We combine $({}^1x_{t+1}^{\mathscr{N}_k}, \dots, {}^rx_{t+1}^{\mathscr{N}_k})$ with the message $m_{t+1}^{\bar{\mathscr{N}_k}} = ({}^1x_{t+1}^{\bar{\mathscr{N}_k}}, \dots, {}^qx_{t+1}^{\bar{\mathscr{N}_k}})$ to form $b_{t+1}^{\mathscr{L}_k} = \{({}^ix_{t+1}^{\mathscr{N}_k}, {}^jx_{t+1}^{\bar{\mathscr{N}_k}}), i = 1, 2, \dots, r;\ j = 1, 2, \dots, q\}$. Thus

$$\begin{aligned} b_{t+1}^{\mathscr{L}_k} &= \phi^k(b_t^{\mathscr{L}_k}, h_{t:t+1}^k) \\ &= \phi^k(b_t^{\mathscr{L}_k}, u_t^{\mathscr{N}_k}, z_t^{\mathscr{N}_k}, y_{t+1}^{\mathscr{N}_k}, m_{t+1}^{\bar{\mathscr{N}_k}}). \end{aligned}$$

With this new approximate information state, each Problem (P_k) is approximated by the following minmax control problem (P'_k).

$$\min_{\gamma'^k \in \Gamma'} \max_{\left\{X_{\mathscr{T}}^{\mathscr{L}_k} \in \mathscr{X}_{\mathscr{T}}^{\mathscr{L}_k, \gamma'^k}\right\}} \left\{ \sum_{t \in \mathscr{T}} \beta^t c^{\mathscr{N}_k}(X_t^{\mathscr{N}_k}, U_t^{\mathscr{N}_k}) \,\middle|\, X_0^{\mathscr{L}_k} = x_0^{\mathscr{L}_k} \right\} \tag{P'_k}$$

$$\text{subject to}\quad X_{t+1}^{\mathscr{N}_k} = \pi^{\mathscr{N}_k}(X_t^{\mathscr{L}_k}, W_t, U_t^{\mathscr{N}_k}) \tag{P'_k-i}$$

$$X_{t+1}^{\bar{\mathscr{N}_k}} \in M_{t+1}^{\bar{\mathscr{N}_k}} \tag{P'_k-ii}$$

$$Y_t^{\mathscr{N}_k} = \theta^{\mathscr{N}_k}(X_t^{\mathscr{N}_k}, W_t) \tag{P'_k-iii}$$

$$Z_t^{\mathscr{N}_k} = \zeta^{\mathscr{N}_k}(X_t^{\mathscr{N}_k}, W_t, U_t^{\mathscr{N}_k}) \tag{P'_k-iv}$$

$$U_t^{\mathscr{N}_k} = \gamma'^k(B_t^{\mathscr{L}_k}) \tag{P'_k-v}$$

where the local states are given by $X_{t+1}^{\mathscr{L}_k} = (X_{t+1}^{\mathscr{N}_k}, X_{t+1}^{\bar{\mathscr{N}_k}})$ and $\Gamma'^k = \{\gamma'^k = (\gamma'^k_1, \gamma'^k_2, \dots, \gamma'^k_T) \,|\, \gamma'^k_t : \mathscr{B}^{\mathscr{L}_k} \to \mathscr{U}^{\mathscr{N}_k}$ for all $t \in \mathscr{T}\}$ represents the set of admissible approximate local defense policies γ'^k.

For finite horizon T we solve Problem (P'_k) backward in time via the following set of recursive equations. Let $b_t^{\mathscr{L}_k} \in \mathscr{B}^{\mathscr{L}_k}$ be the approximate information state at t and $V_t^k(b_t^{\mathscr{L}_k})$ denote the minmax value of Problem (P'_k) from time t on when the approximate information state at t is $b_t^{\mathscr{L}_k}$, $t = 0, 1, \dots, T+1$. Then, for each $b_{T+1}^{\mathscr{L}_k} \in \mathscr{B}^{\mathscr{L}_k}$,

$$V_{T+1}^k(b_{T+1}^{\mathscr{L}_k}) = 0. \tag{4.13}$$

and for $t = 1, 2, \ldots, T$, and each $b_t^{\mathscr{L}_k} \in \mathscr{B}^{\mathscr{L}_k}$,

$$V_t^k(b_t^{\mathscr{L}_k}) = \min_{u_t^{\mathscr{N}_k} \in \mathscr{U}^{\mathscr{N}_k}} \Bigg[\max_{x_t^{\mathscr{L}_k} = (x_t^{\mathscr{N}_k}, x_t^{\bar{\mathscr{N}}_k}) \in b_t^{\mathscr{L}_k}} \Big[c^{\mathscr{N}_k}(x_t^{\mathscr{N}_k}, u_t^{\mathscr{N}_k}) +$$

$$\beta \max_{\substack{w_t \in \mathscr{W}(x_t^{\mathscr{L}_k}) \\ m_{t+1}^{\bar{\mathscr{N}}_k} \in \mathscr{P}(\mathscr{X}^{\bar{\mathscr{N}}_k})}} \Big[\max_{x_{t+1}^{\bar{\mathscr{N}}_k} \in m_{t+1}^{\bar{\mathscr{N}}_k}} \Big[\max_{w_{t+1} \in \mathscr{W}\left(\pi^{\mathscr{N}_k}(x_t^{\mathscr{L}_k}, w_t, u_t^{\mathscr{N}_k}), x_{t+1}^{\bar{\mathscr{N}}_k}\right)}$$

$$V_{t+1}^k\Big(\phi\left(b_t^{\mathscr{L}_k}, u_t^{\mathscr{N}_k}, \zeta^{\mathscr{N}_k}\left(x_t^{\mathscr{N}_k}, w_t, u_t^{\mathscr{N}_k}\right), \theta^{\mathscr{N}_k}\left(\pi^{\mathscr{N}_k}(x_t^{\mathscr{L}_k}, w_t, u_t^{\mathscr{N}_k}, w_{t+1}), m_{t+1}^{\bar{\mathscr{N}}_k}\right)\right)\Big]\Big]\Big]\Bigg] \tag{4.14}$$

where $\mathscr{W}(x_t^{\mathscr{L}_k})$ is defined as the set of events possible from any state x such that the elements $\mathscr{L}_k$ of the state are equal to $x_t^{\mathscr{L}_k}$.

For the infinite horizon case, we solve Problem (P_k') via the set of equations

$$V^k(b^{\mathscr{L}_k}) = \min_{u^{\mathscr{N}_k} \in \mathscr{U}^{\mathscr{N}_k}} \Bigg[\max_{x^{\mathscr{L}_k} = (x^{\mathscr{N}_k}, x^{\bar{\mathscr{N}}_k}) \in b^{\mathscr{L}_k}} \Big[c^{\mathscr{N}_k}(x^{\mathscr{N}_k}, u^{\mathscr{N}_k}) +$$

$$\beta \max_{\substack{w \in \mathscr{W}(x^{\mathscr{L}_k}) \\ m^{\bar{\mathscr{N}}_k} \in \mathscr{P}(\mathscr{X}^{\bar{\mathscr{N}}_k})}} \Big[\max_{x^{\bar{\mathscr{N}}_k} \in m^{\bar{\mathscr{N}}_k}} \Big[\max_{w' \in \mathscr{W}\left(\pi^{\mathscr{N}_k}(x^{\mathscr{L}_k}, w, u^{\mathscr{N}_k}), x^{\bar{\mathscr{N}}_k}\right)}$$

$$V^k\Big(\phi\left(b^{\mathscr{L}_k}, u^{\mathscr{N}_k}, \zeta^{\mathscr{N}_k}\left(x^{\mathscr{N}_k}, w, u^{\mathscr{N}_k}\right), \theta^{\mathscr{N}_k}\left(\pi^{\mathscr{N}_k}(x^{\mathscr{L}_k}, w, u^{\mathscr{N}_k}, w'), m^{\bar{\mathscr{N}}_k}\right)\right)\Big]\Big]\Big]\Bigg] \tag{4.15}$$

for all $b^{\mathscr{L}_k} \in \mathscr{B}^{\mathscr{L}_k}$.

Solving the above recursive equations (Eqs. (4.13) and (4.14) for the finite horizon case, or Eq. (4.15) for the infinite horizon case) for each $k \in \mathscr{K}$ yields a set of (suboptimal) local defense policies $\{\gamma'^1, \gamma'^2, \ldots, \gamma'^{n_k}\}$ for Problem (P).

4.4.3 Scalability

Our approach to the solution of Problem (P) consists of two main steps: (i) The partition of the influence graph into clusters and the formulation of approximate local defense problems (P_k'); and, (ii) The solution of each problem (P_k'). This approach can provide a (suboptimal) solution to the defense problem (P) associated with networks of arbitrarily large size, as we explain below. Suppose that the designer of the defense policy knows its (limited) computational capability. To implement our approach the designer must be able to solve the problems associated with the above described steps.

Forming the influence graph requires the computation of all of the edge weights, ξ_e, as described in Section 4.4.1.2; the complexity of such a computation is of the order of $\mathscr{O}(\sum_{j,i\in\mathscr{N},j\neq i}|\mathscr{X}\times\mathscr{X}^j\times\mathscr{W}\times\mathscr{U}|)=\mathscr{O}(n^2n_wn_u)$. Creating clusters requires the use of the min-cut algorithm, the complexity of which is of the order $\mathscr{O}((n_k-1)n^2)$, where n_k is the number of clusters. The computational complexity of each of the Problems (P'_k) is of the order $\mathscr{O}(|\mathscr{P}(\mathscr{X}^{\mathscr{L}_k})\times\mathscr{Y}^{\mathscr{N}_k}\times\mathscr{Z}^{\mathscr{N}_k}\times\mathscr{U}^{\mathscr{N}_k}\times\mathscr{W}|)=\mathscr{O}(2^{|\mathscr{L}_k|}n_w\prod_{i\in\mathscr{N}_k}(n_y^in_z^in_u^i))$. The problems ($\mathrm{P}'_k$) can be solved in parallel. The above arguments show that the computational complexity associated with the solution of each Problem (P'_k) is the main bottleneck in the application of our approach to the solution of Problem (P). If the computational complexity of each of the problems (P'_k), $k\in\mathscr{K}$, does not exceed the designer's computational capability then our approach can be used to provide a suboptimal solution to Problem (P).

From the above discussion it is clear that an increase in number n_k of clusters will on one hand decrease the computational complexity of each (P'_k), as the dimensionality of each $\mathscr{X}^{\mathscr{N}_k},\mathscr{Y}^{\mathscr{N}_k},\mathscr{Z}^{\mathscr{N}_k},\mathscr{W},\mathscr{U}^{\mathscr{N}_k}$ will decrease, but on the other hand, will decrease the accuracy of the solution of Problem (P) and increase the complexity of the min-cut algorithm. Therefore, in the application of our approach to Problem (P) one has to explore the above described tradeoff between computational complexity and solution quality so as to end up with the best approximation that is compatible with the defender's computational capabilities.

Depending on the structure of the influence graph, modifications to the approach can be taken. In some problems, the influence graph may exhibit some sparsity; in this case the computational complexity associated with clustering can be reduced using spectral clustering with non-backtracking matrix Krzakala et al. (2013).[6] In situations where the influence graph is densely connected, one can use approximations, in addition to those described in this chapter, so as to end up with a scalable approximation to Problem (P). We briefly describe such approximations in the conclusion of the chapter (Section 4.6).

4.5 Example

Consider a system of five hosts. Each host can be in one of four security states (a measure of its security level), ranging from the most secure state, s_1, to the least secure state, s_4. At each time-step, the attacker (nature) can choose to attack the hosts through selection of various attack actions. The attacker is assumed to have access to three types of attack actions: a *null* action, corresponding to not attacking the host; *probe* actions, which increment the security state of the attacked host; and *spread* actions, which allow the attacker to use a host in a degraded security state to attack another host. Following the attack action (nature's event), the defender selects its defense action. The

[6] The spectral clustering with non-backtracking matrix algorithm has lower complexity than the min-cut algorithm for super-sparse graphs.

defender has access to three types of defense actions: a *null* action, corresponding to not specifying any defense action; a *sense* action, which, if invoked on a host, reveals the true security state of the host to the defender; and a *re-image* action, which resets the security state of the host to s_1.

The five host system described above can be modeled using the security model of Section 4.2. Formally, using the notation of our model, each host corresponds to a state element, that is, $\mathscr{N} = \{1,2,3,4,5\}$. The state-space of each element reflects the possible security states that each host can be in, that is, $\mathscr{X}^i = \{s_1, s_2, s_3, s_4\}$. The state-space of the problem is $\mathscr{X} = \prod_{i \in \mathscr{N}} \mathscr{X}^i$. The set of attack actions $\mathscr{W}$ is assumed to decompose into attacks on each host, that is, $\mathscr{W} = \prod_{i \in \mathscr{N}} \mathscr{W}^i$.[7] Each set $\mathscr{W}^i$ consists of attacks $\mathscr{W}^i = \{w^i_\varnothing, w^i_{p_1}, w^i_{p_2}, w^i_{p_3}, w^{ji}_s, w^{ji}_{s'}\}$ on host i, where $w^i_\varnothing$ represents no attack on host i, $w^i_{p_k}$ represents a probe action, incrementing the security state of host i from $x^i_t = s_k$ to $x^i_{t+1} = s_{k+1}$, and both w^{ji}_s and $w^{ji}_{s'}$ represent spread actions, allowing the attacker to use another host j if it is in state $x^j_t = s_4$ to attack host i. Specifically, w^{ji}_s brings the state of host i from $x^i_t = s_1$ to $x^i_{t+1} = s_3$ and $w^{ji}_{s'}$ brings the state of host i from $x^i_t = s_2$ to $x^i_{t+1} = s_3$. To make the example more interesting (resulting in a more diverse set of weights in the influence graph), we assume that the attacker has limited spreading capabilities, that is, $\mathscr{W}^1$ contains the spreading actions $\{w^{5,1}_s, w^{5,1}_{s'}\}$, $\mathscr{W}^2$ contains $\{w^{1,2}_s, w^{1,2}_{s'}, w^{3,2}_s\}$, $\mathscr{W}^3$ contains $\{w^{1,3}_s, w^{4,3}_s\}$, $\mathscr{W}^4$ contains $\{w^{3,4}_s\}$, and $\mathscr{W}^5$ contains $\{w^{1,5}_s\}$. The set of defense actions $\mathscr{U} = \prod_{i \in \mathscr{N}} \mathscr{U}^i$ is described in terms of the action-space of each element. Specifically, $\mathscr{U}^i = \{u^i_\varnothing, u^i_s, u^i_r\}$, where $u^i_\varnothing$ represents the defender not taking any action on host i, u^i_s represents the sense action on host i, and u^i_r represents the re-image action on host i. Neither the null actions or the sense action have an affect on the evolution of the state.

In terms of the state update function of Eq. (4.6), the evolution of the each state element can be written as follows

$$x^i_{t+1} = \pi^i(x^i_t, w_t, u^i_t) = \begin{cases} s_1 & \text{if } u^i_t = u^i_r \\ s_3 & \text{else if } (w^{ji}_s \in w_t,\, x^j_t = s_4,\, x^i_t = s_1) \text{ or} \\ & \qquad\qquad (w^{ji}_{s'} \in w_t,\, x^j_t = s_4,\, x^i_t = s_2) \\ s_{k+1} & \text{else if } w^i_t = w^i_{p_k},\, x^i_t = s_k,\, k = 1,2,3 \\ x^i_t & \text{otherwise.} \end{cases}$$

We assume that the defender is only able to observe the spreading actions, but cannot observe the

[7]Note that in our model of Section 4.2, the attacks are not necessarily decomposable into attacks on each element; however, for the purposes of our example, we assume (for simplicity) that the system-wide attack can be described as the collection of attacks on each element.

attacker's null or probe actions. Formally,

$$y_t^i = \theta^i(x_t^i, w_t) = \begin{cases} w \text{ if for any } w = w_s^{ji} \text{ or } w = w_{s'}^{ji} \text{ in } w_t \\ \varnothing \text{ otherwise.} \end{cases}$$

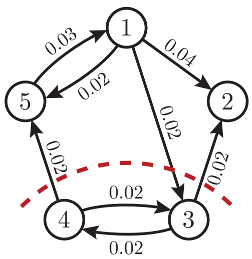
The defender's action observations are

$$z_t^i = \zeta^i(x_t^i, u_t^i) = \begin{cases} s_1 \text{ if } u_t^i = u_r^i \\ x_t^i \text{ else if } u_t^i = u_s^i \\ \varnothing \text{ otherwise.} \end{cases}$$

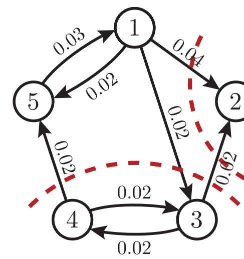
Finally, the (instantaneous) cost that the defender incurs at time t is simply $c(x_t, u_t) = c^1(x_t^1, u_t^1) + \cdots + c^5(x_t^5, u_t^5)$ where $x_t = (x_t^1, x_t^2, \ldots, x_t^5)$ and $u_t = (u_t^1, u_t^2, \ldots, u_t^5)$. The defense problem (P) can now be written. In previous work Rasouli et al. (2014), we were able to obtain a defense policy for a similar problem with $n = 3$ elements using the approximate information state (the set of states compatible with the defender's information at t) described in Section 4.4.2. For larger problems (for instance, the $n = 5$ problem just described), we must employ the decomposition approach proposed in this chapter to permit computation of approximate defense policies.

The influence graph for the example problem can now be constructed. Assume that the computational capability of the defender is such that it can solve Problem (P) for systems consisting of $n^{\max} = 3$ or fewer elements.[8] In the current example $n = 5 > 3 = n^{\max}$, so we must decompose the problem into local defense problems and determine local defense policies. As described in Section 4.4.1.2, the construction of the influence graph is performed by analyzing the functions and the sets of actions of both the defender and attacker with the weights computed according to Eq. (4.11). For illustration purposes, we show how to compute one of the weights, specifically $\xi_{5,1}$. The remaining edge weights are calculated in a similar fashion. To calculate $\xi_{5,1}$, we need to count the cases where the state update of element $i = 1$ functionally depends on the state of element $i = 5$, as described by Definition 4.4.1; we enumerate over all values of x^1, u^1, w^1, x^5, $\hat{x}^5 \neq x^5$, u^5, w^5, $x^{-(1,5)}$, and $w^{-(1,5)}$. The total number of different values which is the normalization term in Eq. (4.11), is $d = (\prod_{l \in \mathcal{N} \setminus \{i,j\}} n_x^l) n_x^i n_x^j (n_x^j - 1) n_w n_u^i = (4^3) \cdot 4 \cdot 4 \cdot (4-1) \cdot (6 \cdot 7 \cdot 6 \cdot 5 \cdot 5) \cdot 3 = 58060800$. Only in those cases that all of the following conditions are satisfied, the inequality in Eq. 4.11 holds: (i) A spread attack is launched from element $j = 5$ to $i = 1$ while the state of element $i = 1$ is such that the spread attack is effective, i.e. $(w^5 = w_s^{5,1}, x^1 = s^1)$ or $(w^5 = w_{s'}^{5,1}, x^1 = s^2)$. (ii) the state of element $j = 5$ allows the the attack to be launched in x_5 but does not allow it in $\hat{x}_5$ or vice versa i.e. $(x_5 = s_4, \hat{x}^5 \in \{s_1, s_2, s_3\})$ or $(x_5 \in \{s_1, s_2, s_3\}, \hat{x}^5 = s_4)$. (iii) There are no similar

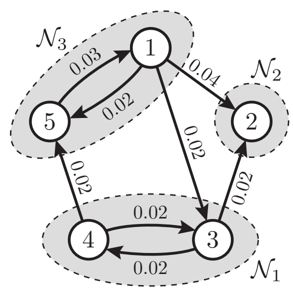
[8] The quantity $n^{\max}$ is determined by taking into account the defender's computational capability and the computational complexity of Problem (P_k'), as described in Section 4.4.3.



(a) First cut.



(b) Second cut.



(c) Final set of clusters.

Figure 4.4: Repeated application of the mincut algorithm to obtain three clusters, $\mathscr{N}_1 = \{3,4\}$, $\mathscr{N}_2 = \{2\}$, and $\mathscr{N}_3 = \{1,5\}$.

effective spread attacks on element $i = 1$ from other elements $k \in \{2,3,4\}$; since there exists no spread attack from elements $k \in \{2,3,4\}$ to element $i = 1$, their state and attack events can take any possible value i.e. $x^{-(1,5)} \in \mathscr{X}^{-(1,5)}$, and $w^{-(1,5)} \in \mathscr{W}^{-(1,5)}$. (iv) The defender does not deploy a defense action that nullifies the effect of the spread attack from element $j = 5$ to element $i = 1$, that is, $u^1 \in \{u^1_{\varnothing}, u^1_s\}$. The above conditions ensure the updated state of element $i = 1$ differs between the states x^5 and $\hat{x}^5$. The total number of such cases is $2 \cdot 2 \cdot (1 \cdot 3 + 3 \cdot 1) \cdot 4^3 \cdot (7 \cdot 6 \cdot 5 \cdot 5) = 1612800$. The normalized weight, computed using Eq. (4.11), is thus $\xi_{5,1} = 1612800/58060800 \approx 0.03$. The complete influence graph is depicted in Fig. 4.4.

Using the edge weights in the influence graph, the min-cut algorithm Johnson et al. (1993) can now be applied in order to partition (cluster) the graph. The first application of the min-cut algorithm, shown in Fig. 4.4(a), results in the clusters $\{1,2,5\}$ and $\{3,4\}$. Notice that under this clustering, the existence of edges $(3,2)$ and $(4,5)$ would result in one of the local defense problems containing all 5 elements in its local state-space, violating the $n^{\max} = 3$ limit. As a result, we apply the min-cut algorithm once more, as shown in Fig. 4.4(b). With the new cut we can see that the resulting set of clusters, $\mathscr{N}_1 = \{3,4\}$, $\mathscr{N}_2 = \{2\}$, and $\mathscr{N}_3 = \{1,5\}$, shown in Fig. 4.4(c), satisfy the $n^{\max}$ limit. This can be seen by writing the set of state indices for the local defense problems, $\mathscr{L}_1 = \{1,3,4\}$, $\mathscr{L}_2 = \{1,2,3\}$, and $\mathscr{L}_3 = \{1,4,5\}$ and noticing that $|\mathscr{L}_1| = |\mathscr{L}_2| = |\mathscr{L}_3| = 3 = n^{\max}$.

Using the clusters of Fig. 4.4(c), the corresponding local defense problems can be defined. The internal state spaces of each problem are defined by the set of nodes within the corresponding cluster, that is, $\mathscr{X}^{\mathscr{N}_1} = \mathscr{X}^3 \times \mathscr{X}^4$, $\mathscr{X}^{\mathscr{N}_2} = \mathscr{X}^2$, and $\mathscr{X}^{\mathscr{N}_3} = \mathscr{X}^1 \times \mathscr{X}^5$. The corresponding state update functions are

$$\begin{aligned}
\pi_{t+1}^{\mathscr{N}_1} &= \pi^{\mathscr{N}_1}(x_t^{\mathscr{L}_1}, w_t, u_t^{\mathscr{N}_1}) \\
&= \big(\pi^3(x_t^{\mathscr{L}_1}, w_t, u_t^3), \pi^4(x_t^{\mathscr{L}_1}, w_t, u_t^4)\big) \\
\pi_{t+1}^{\mathscr{N}_2} &= \pi^{\mathscr{N}_2}(x_t^{\mathscr{L}_2}, w_t, u_t^{\mathscr{N}_2}) \\
&= \pi^2(x_t^{\mathscr{L}_2}, w_t, u_t^2) \\
\pi_{t+1}^{\mathscr{N}_3} &= \pi^{\mathscr{N}_3}(x_t^{\mathscr{L}_3}, w_t, u_t^{\mathscr{N}_3}) \\
&= \big(\pi^1(x_t^{\mathscr{L}_3}, w_t, u_t^1), \pi^5(x_t^{\mathscr{L}_3}, w_t, u_t^5)\big)
\end{aligned}$$

where the local states are $x_t^{\mathscr{L}_1} = \{x_t^1, x_t^3, x_t^4\}$, $x_t^{\mathscr{L}_2} = \{x_t^1, x_t^2, x_t^3\}$, and $x_t^{\mathscr{L}_3} = \{x_t^1, x_t^4, x_t^5\}$ and the internal actions are $u_t^{\mathscr{N}_1} \in \mathscr{U}^{\mathscr{N}_1} = \mathscr{U}^3 \times \mathscr{U}^4$, $u_t^{\mathscr{N}_2} \in \mathscr{U}^{\mathscr{N}_1} = \mathscr{U}^2$, and $\mathscr{U}^{\mathscr{N}_3} = \mathscr{U}^1 \times \mathscr{U}^5$. The cost functions of each local defense problem are $c^{\mathscr{N}_1}(x_t^{\mathscr{N}_1}, u_t^{\mathscr{N}_1}) = c^3(x_t^3, u_t^3) + c^4(x_t^4, u_t^4)$, $c^{\mathscr{N}_2}(x_t^{\mathscr{N}_2}, u_t^{\mathscr{N}_2}) = c^2(x_t^2, u_t^2)$, and $c^{\mathscr{N}_2}(x_t^{\mathscr{N}_2}, u_t^{\mathscr{N}_2}) = c^1(x_t^1, u_t^1) + c^5(x_t^5, u_t^5)$. Similarly, each local defense problem's observations are $y_t^{\mathscr{N}_1} = (y_t^3, y_t^4)$, $z_t^{\mathscr{N}_1} = (z_t^3, z_t^4)$, $y_t^{\mathscr{N}_2} = y_t^2$, $z_t^{\mathscr{N}_2} = z_t^2$, and $y_t^{\mathscr{N}_3} = (y_t^1, y_t^5)$, $z_t^{\mathscr{N}_3} = (z_t^1, z_t^5)$. The message sent from local defense problem l to local defense problem k at time t, denoted by m_t^{kl}, belongs to the powerset of states that can influence the evolution of elements in the internal state space $\mathscr{N}_k$. Specifically, for the example, local defense problem $k = 1$ receives a message from problem $k = 3$, described by $m_t^{13} \in \mathscr{P}(\mathscr{X}^{\bar{\mathscr{N}}_{13}}) = \mathscr{P}(\mathscr{X}^1)$ (representing all possible states the element $i = 1$ can be in). Similarly, local defense problem $k = 2$ receives two messages, one from problem $k = 1$, $m_t^{21} \in \mathscr{P}(\mathscr{X}^{\bar{\mathscr{N}}_{21}}) = \mathscr{P}(\mathscr{X}^3)$, and one from problem $k = 3$, $m_t^{23} \in \mathscr{P}(\mathscr{X}^{\bar{\mathscr{N}}_{23}}) = \mathscr{P}(\mathscr{X}^1)$ (these messages can be summarized by the aggregate message $m_t^{\mathscr{N}_2} \in \mathscr{P}(\mathscr{X}^3) \times \mathscr{P}(\mathscr{X}^1)$). Lastly, local defense problem $k = 3$ receives a message from problem $k = 1$, $m_t^{31} \in \mathscr{P}(\mathscr{X}^{\bar{\mathscr{N}}_{31}}) = \mathscr{P}(\mathscr{X}^4)$. No other messages are exchanged. The approximate information state spaces for the local defense problems are $\mathscr{B}^{\mathscr{L}_1} = \mathscr{P}(\mathscr{X}^1 \times \mathscr{X}^3 \times \mathscr{X}^4)$, $\mathscr{B}^{\mathscr{L}_2} = \mathscr{P}(\mathscr{X}^1 \times \mathscr{X}^2 \times \mathscr{X}^3)$ and $\mathscr{B}^{\mathscr{L}_3} = \mathscr{P}(\mathscr{X}^1 \times \mathscr{X}^4 \times \mathscr{X}^5)$. The dynamic programs (for the finite horizon defense problem) can be written in a similar fashion to Eq. (4.13) and (4.14). The equations can be solved using standard value iteration-type algorithms to obtain the local defense policies γ'^k, $k \in \mathscr{K}$.

4.6 Conclusion and Future Directions

We studied a cyber-security problem from the defender's point of view. This is a control problem where the defender's goal is to determine a defense policy to protect the system against an attacker that is modeled by nature. The system's security status evolves dynamically over time; its evolution depends on the defender's actions and the attack events. The defender has imperfect information about the system's security status and takes a conservative approach to the system's defense. Specifically, the defender's goal is to minimize the worst possible damage to the system caused by attack (nature's) events. Therefore, the defender has to solve a minmax control problem with imperfect observation so as to determine an optimal defense policy.

The defender's imperfect observation combined with the high dimensionality of the system's state and the minmax objective result in a complicated information state that renders the computation of an optimal defense policy intractable, necessitating approximations. The approximation we present is based on decomposing the defense problem into local defense problems and solving for local defense policies. We form local defense problems by first forming the influence graph – a weighted directed graph quantifying the dependencies between the system's elements. Next, we cluster the influence graph into clusters of strongly-dependent elements using the min-cut algorithm. Using the clusters and the dependencies among them, we form a local defense problem for each cluster. The control of each local defense problem is further approximated by focusing on an approximate information state that allows the computation of a policy sequentially using dynamic programming ideas.

Our approach has two computational requirements: (i) forming the influence graph and clusters; (ii) computing a defense policy for the local defense problem associated with each cluster. Given the defender's computational capabilities, we can address these requirements irrespectively of the system's size (the dimensionality of the system's state) as follows: we can form a large number of clusters so that the size of each local defense problem is compatible with the defender's computational capability. Consequently, the approach to the dynamic defense problem of cyber networks described in this chapter is scalable.

For some instances of our security model, the resulting influence graph may not permit a partitioning into clusters that satisfy the designer's computational capability. For example, consider a completely connected influence graph. In such a graph, each local state space $\mathscr{X}^{\mathscr{L}_k}$ would be equal to the complete state space $\mathscr{X}$. In such a situation, no clustering that satisfies the constraint would exist and we would not be able to compute local defense policies. However, we believe that, for practical purposes, influence graphs will have at least some sparsity that can be exploited, allowing our decomposition to be applied to obtain a suboptimal defense policy. In the rare event that the influence graph doesn't permit the required clustering, one can employ alternate (more aggressive) approximation techniques that summarize the available information (for example, collapsing all exogenous elements into a single worst-case element).

In the cyber security model studied in this chapter the attacker's behavior is fixed in advance and modeled by the (state-dependent) events that occur in nature. The situation where both the defender and the attacker are strategic and have different objectives is not captured by the model of Section 4.2. Such a situation gives rise to a dynamic game with asymmetric information. Preliminary results on such games can be found in Ouyang et al. (2017); Tavafoghi et al. (2016).

CHAPTER 5

Conclusion and Future Directions

This thesis presented cyber-physical systems design (CPS Design). CPS are systems that have both physical and cyber components. Examples are electricity networks, security networks, transportation networks, smart connected medical devices, etc. CPS are growing in our modern life because of the technological developments in embedded systems and smart devices.

CPS design is complicated and requires interdisciplinary studies of economics, engineering and computer science because of the distinguishing features of these systems which are the presence of strategic decision makers, the large-scale nature of CPS, and the existence of complex constraints. These distinguishing features of the CPS require a new approach for their design. Strategic users cannot be controlled directly, rather they should be controlled by providing appropriate incentives. Therefore, CPS design requires mechanism (such as market and auction) design. On the other hand the large scale nature of the CPS means these systems cannot be controlled optimally as the determination of the optimal control strategy is computationally infeasible; instead the designer should look for approximate algorithms and mechanisms that allow for the determination of good suboptimal but feasible control. These approximations can be made based on heuristics. Finally, the complex constraints imposed by real world (such as physical or policy constraints) introduce new models and design problems which are application specific. All these features mean that a new system design approach for the CPS design should be built. Note that, in addition to their distinguishing features, CPS have the features of other engineering systems such as presence of uncertainty and dynamic interaction of agents.

Networks are a central tool for CPS design. Networks allow for modeling a lot of the the complex constraints in CPS. Furthermore, their structure/topology can be used for decomposition of large scale CPS into small subsystems.

In this thesis, we studied CPS design by focusing on two specific problems, designing markets with complex constraints, and decomposition of large-scale CPS. For the first problem we focused on electricity markets and for the second one we considered network security.

We studied the design of markets with complex constraints in two stages: electricity markets with complex policy constraints, and with complex physical constraints. For markets with policy constraints, we focused on implementing electricity policy targets i.e. sustainability, reliability and price efficiency by designing markets that are also budget balanced, individually rational and social welfare maximizing. To achieve our goal, we developed a framework for designing efficient auctions with constraints including individual and homogeneous (additive) joint constraints. Our results, shed light on major policy debates including price/offer caps, capacity-and-energy solutions vs energy-only solutions, and carbon market vs carbon tax. For markets with physical constraints, we focused on designing efficient spot markets when complex network constraints due to Kirchhoff's laws exist. The market should also be individually rational, price efficient, budget balanced and social welfare maximizing. To achieve this goal, we used ideas from local public goods where the network is considered a public good whose use should be determined by all users collectively.

Second, we studied decomposition of large-scale CPS, motivated by network security issues. We introduced the notion of influence graph that captures the connectivity of the elements in a network. The influence graph then allows for decomposition of the system into subsystems by clustering connected elements together. Each cluster is controlled by a local defense algorithm which can exchange information with other local defense algorithms. This approach allows for heuristic and suboptimal defense of the network which is computationally feasible.

Our research can be extended in two directions. The first direction is designing other CPS such as transportation networks, water networks, or gas networks by modeling their constraints. The second direction is to add other features of CPS especially the information asymmetry among strategic agents. When the strategic agents have asymmetric information, the CPS design requires implementation in perfect Bayesian equilibria or dominant strategy implementation of the social choice function instead of the Nash implementation approach used in this thesis.

As fast as CPS are growing in our life, CPS design will be more and more critical and beneficial to the well-being of the human society.

BIBLIOGRAPHY

World energy outlook 2016. Technical report, International Energy Agency, 2016.

International energy outlook 2016. Technical report, U.S. Energy Information Administration, 2016.

Daron Acemoglu, Kostas Bimpikis, and Asuman Ozdaglar. Price and capacity competition. *Games and Economic Behavior*, 66(1):1–26, 2009.

Daron Acemoglu, Ufuk Akcigit, Douglas Hanley, and William Kerr. Transition to clean technology. Technical report, National Bureau of Economic Research, 2014.

Daron Acemoglu, Ali Kakhbod, and Asuman Ozdaglar. Competition in electricity markets with renewable energy sources. 2016a.

Daron Acemoglu, Asuman Ozdaglar, and Insoon Yang. A dynamic industrial equilibrium model for transition to cleaner technologies. 2016b.

Sam Aflaki and Serguei Netessine. Strategic investment in renewable energy sources: The effect of supply intermittency. 2016.

Mark Ahlstrom, Erik Ela, Jenny Riesz, Jonathan O'Sullivan, Benjamin F Hobbs, Mark O'Malley, Michael Milligan, Paul Sotkiewicz, and Jim Caldwell. The evolution of the market: Designing a market for high levels of variable generation. *Power and Energy Magazine, IEEE*, 13(6):60–66, 2015.

René Aïd, Gilles Chemla, Arnaud Porchet, and Nizar Touzi. Hedging and vertical integration in electricity markets. *Management Science*, 57(8):1438–1452, 2011.

Mohammad Akbarpour and Afshin Nikzad. Approximate random allocation mechanisms. 2017.

Mohammad Akbarpour, Shengwu Li, and Shayan Oveis Gharan. Thickness and information in dynamic matching markets. 2017.

Saed Alizamir, Francis de Véricourt, and Peng Sun. Efficient feed-in-tariff policies for renewable energy technologies. *Operations Research*, 64(1):52–66, 2016.

Blaise Allaz and Jean-Luc Vila. Cournot competition, forward markets and efficiency. *Journal of Economic theory*, 59(1):1–16, 1993.

Tansu Alpcan and Tamer Başar. *Network security: A decision and game-theoretic approach*. Cambridge University Press, 2010.

Edward J Anderson. On the existence of supply function equilibria. *Mathematical Programming*, 140(2):323–349, 2013.

Edward J Anderson and Xinmin Hu. Finding supply function equilibria with asymmetric firms. *Operations Research*, 56(3):697–711, 2008.

Ross Baldick. Wind and energy markets: A case study of texas. *Systems Journal, IEEE*, 6(1):27–34, 2012.

Ross Baldick and William Hogan. Polynomial approximations and supply function equilibrium stability. *Center for Business and Government, Harvard University, Cambridge, Massachusetts*, 2138, 2004.

Ross Baldick and William W Hogan. Stability of supply function equilibria implications for daily versus hourly bids in a poolco market. *Journal of Regulatory Economics*, 30(2):119–139, 2006.

Ross Baldick, William W Hogan, et al. *Capacity constrained supply function equilibrium models of electricity markets: stability, non-decreasing constraints, and function space iterations*. University of California Energy Institute, 2001.

John S Baras and Matthew R James. Robust and risk-sensitive output feedback control for finite state machines and hidden markov models. Technical report, 1994.

Dirk Bergemann and Juuso Välimäki. Information acquisition and efficient mechanism design. *Econometrica*, 70(3):1007–1033, 2002.

Pierre Bernhard. Expected values, feared values, and partial information optimal control. In *New Trends in Dynamic Games and Applications*, pages 3–24. Springer, 1995.

Pierre Bernhard. Max-plus algebra and mathematical fear in dynamic optimization. *Set-Valued Analysis*, 8(1-2):71–84, 2000.

Pierre Bernhard. Minimax – or feared value – $L1/L\infty$ control. *Theoretical Computer Science*, 293 (1):25–44, 2003.

D Bertsekas and I Rhodes. Sufficiently informative functions and the minimax feedback control of uncertain dynamic systems. *IEEE Transactions on Automatic Control*, 18(2):117–124, 1973.

Dimitri P Bertsekas. Control of uncertain systems with a set-membership description of the uncertainty. Technical report, DTIC Document, 1971.

Dimitri P Bertsekas and Ian B Rhodes. On the minimax feedback control of uncertain dynamic systems. In *Decision and Control, 1971 IEEE Conference on*, volume 10, pages 451–455. IEEE, 1971.

Miles Bidwell. Reliability options: a market-oriented approach to long-term adequacy. *The Electricity Journal*, 18(5):11–25, 2005.

Kostas Bimpikis, Shayan Ehsani, and Rahmi Ilkilic. Cournot competition in networked markets. In *EC*, page 733, 2014.

Mario Blazquez de Paz. Designing electricity auctions in the presence of transmission constraints. 2014.

Roger E Bohn, Michael C Caramanis, and Fred C Schweppe. Optimal pricing in electrical networks over space and time. *The Rand Journal of Economics*, pages 360–376, 1984.

Severin Borenstein. The trouble with electricity markets: understanding california's restructuring disaster. *The Journal of Economic Perspectives*, 16(1):191–211, 2002.

Severin Borenstein. The private and public economics of renewable electricity generation. *The Journal of Economic Perspectives*, 26(1):67–92, 2012.

Severin Borenstein, James B Bushnell, and Frank A Wolak. Measuring market inefficiencies in california's restructured wholesale electricity market. *American Economic Review*, pages 1376–1405, 2002.

Severin Borenstein, James Bushnell, Frank A Wolak, and Matthew Zaragoza-Watkins. Expecting the unexpected: Emissions uncertainty and environmental market design. Technical report, National Bureau of Economic Research, 2015.

Tilman Börgers. *An Introduction to the Theory of Mechanism Design*. Oxford University Press, USA, 2015.

Jim B Bushnell and Jun Ishii. An equilibrium model of investment in restructured electricity markets. *Center for the Study of Energy Markets*, 2007.

Lucy Butler and Karsten Neuhoff. Comparison of feed-in tariff, quota and auction mechanisms to support wind power development. *Renewable energy*, 33(8):1854–1867, 2008.

Michael Cadwalader, Paul Gribik, William Hogan, and Susan Pope. Extended lmp and financial transmission rights, 2010.

CAISO. Market surveillance committee of the california iso 2014 final report. 2014.

Judith B Cardell, Carrie Cullen Hitt, and William W Hogan. Market power and strategic interaction in electricity networks. *Resource and Energy Economics*, 19(1):109–137, 1997.

Hung-Po Chao and Stephen Peck. A market mechanism for electric power transmission. *Journal of regulatory economics*, 10(1):25–59, 1996.

Stefano P Coraluppi and Steven I Marcus. Risk-sensitive and minimax control of discrete-time, finite-state markov decision processes. *Automatica*, 35(2):301–309, 1999.

Toby Couture and Yves Gagnon. An analysis of feed-in tariff remuneration models: Implications for renewable energy investment. *Energy policy*, 38(2):955–965, 2010.

Peter Cramton and Steven Stoft. A capacity market that makes sense. *The Electricity Journal*, 18 (7):43–54, 2005.

Peter Cramton, Axel Ockenfels, Steven Stoft, et al. Capacity market fundamentals. *Economics of Energy & Environmental Policy*, 2(2):27–46, 2013.

Carl Davidson and Raymond Deneckere. Long-run competition in capacity, short-run competition in price, and the cournot model. *The Rand Journal of Economics*, pages 404–415, 1986.

María-Ángeles de Frutos and Natalia Fabra. Endogenous capacities and price competition: The role of demand uncertainty. *International Journal of Industrial Organization*, 29(4):399–411, 2011.

Avinash Dixit. Irreversible investment with price ceilings. *Journal of Political Economy*, pages 541–557, 1991.

Anthony Downward, Golbon Zakeri, and AB Philpottt. On cournot equilibria in electricity transmission networks. *Operations research*, 58(4-part-2):1194–1209, 2010.

David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

Andreas Ehrenmann and Karsten Neuhoff. A comparison of electricity market designs in networks. *Operations research*, 57(2):274–286, 2009.

Wedad Elmaghraby and Shmuel S Oren. The efficiency of multi-unit electricity auctions. *The Energy Journal*, pages 89–116, 1999.

Euroelectric. Integrating intermittent renewables sources into the eu electricity system by 2020: challenges and solutions. *http://www.eurelectric.org/*, 2010.

Natalia Fabra, Nils-Henrik Fehr, and David Harbord. Designing electricity auctions. *The RAND Journal of Economics*, 37(1):23–46, 2006.

FERC. Ferc docket nos. er10-787-000, el10-50-000 and el10-57-000, first brief of iso new england inc. 2012.

Bingrui Foo, Matthew W Glause, Gaspar M Howard, Yu-Sung Wu, Saurabh Bagchi, and Eugene H Spafford. Intrusion response systems: a survey. *Information Assurance: Dependability and Security in Networked Systems. Morgan Kaufmann, Burlington*, 2008.

Alfredo Garcia and Zhijiang Shen. Equilibrium capacity expansion under stochastic demand growth. *Operations research*, 58(1):30–42, 2010.

Talat S Genc and Stanley S Reynolds. Supply function equilibria with capacity constraints and pivotal suppliers. *International Journal of Industrial Organization*, 29(4):432–442, 2011.

Richard J Green and David M Newbery. Competition in the british electricity spot market. *Journal of political economy*, pages 929–953, 1992.

Arthur Henriot and Jean-Michel Glachant. Melting-pots and salad bowls: The current debate on electricity market design for integration of intermittent res. *Utilities Policy*, 27:57–64, 2013.

Benjamin F Hobbs. Designing and modeling power markets to support optimal decisions. *Workshop on Optimization and Equilibrium in Energy Economics, IPAM, UCLA*, 2015.

William Hogan. Market and electricity restructuring capacity auctions and minimum offer price rules. 2016.

William W Hogan. Contract networks for electric power transmission. *Journal of regulatory economics*, 4(3):211–242, 1992.

William W Hogan. A cleaner energy system: Renewable energy and electricity market design [in my view]. *Power and Energy Magazine, IEEE*, 13(4):112–109, 2015a.

William W Hogan. Electricity markets and the clean power plan. *The Electricity Journal*, 28(9): 9–32, 2015b.

William W Hogan et al. On an energy only electricity market design for resource adequacy. *California ISO*, 2005.

William W Hogan et al. Electricity scarcity pricing through operating reserves. *Economics of Energy & Environmental Policy*, 2(2):65–86, 2013.

Pär Holmberg and David Newbery. The supply function equilibrium and its policy implications for wholesale electricity auctions. *Utilities Policy*, 18(4):209–226, 2010.

Par Holmberg and Frank Wolak. Electricity markets: Designing auctions where suppliers have uncertain costs. 2015.

Leonid Hurwicz. Outcome functions yielding walrasian and lindahl allocations at nash equilibrium points. *The Review of Economic Studies*, 46(2):217–225, 1979.

Zakira Inayat, Abdullah Gani, Nor Badrul Anuar, Muhammad Khurram Khan, and Shahid Anwar. Intrusion response systems: Foundations, design, and challenges. *J. Netw. Comput. Appl.*, 62: 53–74, 2016.

Ellis L Johnson, Anuj Mehrotra, and George L Nemhauser. Min-cut clustering. *Mathematical programming*, 62(1-3):133–151, 1993.

Paul Joskow and Jean Tirole. Retail electricity competition. *The Rand Journal of Economics*, 37(4): 799–815, 2006.

Paul Joskow and Jean Tirole. Reliability and competitive electricity markets. *The Rand Journal of Economics*, 38(1):60–84, 2007.

Paul L Joskow. Competitive electricity markets and investment in new generating capacity. *AEI-Brookings Joint Center Working Paper*, (06-14), 2006.

Paul L Joskow and Jean Tirole. Transmission rights and market power on electric power networks. *The Rand Journal of Economics*, pages 450–487, 2000.

Nathaniel O Keohane. Cap and trade, rehabilitated: Using tradable permits to control us greenhouse gases. *Review of Environmental Economics and policy*, page ren021, 2009.

Jon Kleinberg and Éva Tardos. Algorithm design. 2006.

Paul D Klemperer and Margaret A Meyer. Supply function equilibria in oligopoly under uncertainty. *Econometrica: Journal of the Econometric Society*, pages 1243–1277, 1989.

A Gürhan Kök, Kevin Shang, and Şafak Yücel. Impact of electricity pricing policies on renewable energy investments and carbon emissions. *Management Science*, 2016.

Tarjei Kristiansen. Markets for financial transmission rights. *Energy Studies Review*, 13(1), 2004.

Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

Yonghong Kuang, Yongjun Zhang, Bin Zhou, Canbing Li, Yijia Cao, Lijuan Li, and Long Zeng. A review of renewable energy utilization in islands. *Renewable and Sustainable Energy Reviews*, 59:504–513, 2016.

Panqanamala Ramana Kumar and Pravin Varaiya. *Stochastic Systems: Estimation, Identification and Adaptive Control*. Prentice-Hall, Inc., 1986.

Jay Ligatti, Lujo Bauer, and David Walker. Edit automata: Enforcement mechanisms for run-time security policies. *International Journal of Information Security*, 4(1):2–16, 2005.

Kongwei Lye and Jeannette M Wing. Game strategies in network security. *International Journal of Information Security*, 4(1-2):71–86, 2005.

Mohammad Hossein Manshaei, Quanyan Zhu, Tansu Alpcan, Tamer Başar, and Jean-Pierre Hubaux. Game theory meets network security and privacy. *ACM Computing Surveys (CSUR)*, 45(3):25, 2013.

Gilbert E Metcalf. Designing a carbon tax to reduce us greenhouse gas emissions. *Review of Environmental Economics and Policy*, page ren015, 2008.

Erik Miehling, Mohammad Rasouli, and Demosthenis Teneketzis. Optimal defense policies for partially observable spreading processes on Bayesian attack graphs. In *Proceedings of the Second ACM Workshop on Moving Target Defense*, pages 67–76. ACM, 2015.

Paul Milgrom. *Discovering Prices: Auction Design in Markets with Complex Constraints*. Columbia University Press, 2017.

Paul Milgrom and Ilya Segal. Deferred-acceptance clock auctions and radio spectrum reallocation. Technical report, Working Paper, 2017.

Frederic H Murphy and Yves Smeers. Generation capacity expansion in imperfectly competitive restructured electricity markets. *Operations research*, 53(4):646–661, 2005.

Robert Novy-Marx. An equilibrium model of investment under uncertainty. *Review of Financial Studies*, 20(5):1461–1502, 2007.

Yi Ouyang, Hamidreza Tavafoghi, and Demosthenis Teneketzis. Dynamic games with asymmetric information: Common information based perfect bayesian equilibria and sequential decomposition. *IEEE Transactions on Automatic Control*, 62(1):222–237, 2017.

Georgia Perakis and Wei Sun. Efficiency analysis of cournot competition in service industries with congestion. *Management Science*, 60(11):2684–2700, 2014.

Geoffrey Pritchard, Golbon Zakeri, and Andrew Philpott. A single-settlement, energy-only electric power market for unpredictable and intermittent participants. *Operations research*, 58(4-part-2): 1210–1219, 2010.

Mohammad Rasouli, Erik Miehling, and Demosthenis Teneketzis. A supervisory control approach to dynamic cyber-security. In *Decision and Game Theory for Security*, pages 99–117. Springer, 2014.

Ingmar Ritzenhofen, John R Birge, and Stefan Spinler. Robustness of renewable energy support schemes facing uncertainty and regulatory ambiguity. 2014.

Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

Fabien A Roques and Nicos Savva. Investment under uncertainty with price ceilings in oligopolies. *Journal of Economic Dynamics and Control*, 33(2):507–524, 2009.

David Salant and Robert Stoddard. Auction design for capacity markets. 2008.

Fred B. Schneider. Enforceable security policies. *ACM Trans. Inf. Syst. Secur.*, 3(1):30–50, February 2000.

Nicola Secomandi and Sunder Kekre. Optimal energy procurement in spot and forward markets. *Manufacturing & Service Operations Management*, 16(2):270–282, 2014.

Alireza Shameli-Sendi, Naser Ezzati-Jivan, Masoume Jabbarifar, and Michel Dagenais. Intrusion response systems: survey and taxonomy. *Int. J. Comput. Sci. Netw. Secur*, 12(1):1–14, 2012.

Fernando J. de Sisternes and John E. Parsons. The impact of uncertainty on the need and design of capacity remuneration mechanisms in low-carbon power systems. *http://web.mit.edu/ferds/www/2016-004.pdf*, 2016.

Yi Sun, Felix F Wu, and Yunhe Hou. Financing long-term generation capacity in a reference price oriented capacity market. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pages 1–7. IEEE, 2010.

Hamidreza Tavafoghi, Yi Ouyang, and Demosthenis Teneketzis. On stochastic dynamic games with delayed sharing information structure. In *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pages 7002–7009. IEEE, 2016.

UNDP. Sustainable development goals, goal 7: Affordable and clean energy, 2015. URL http://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-7-affordable-and-clean-energy.html.

Rakesh V Vohra. *Mechanism Design: a Linear Programming Approach*, volume 47. Cambridge University Press, 2011.

James JD Wang and Jaime F Zender. Auctioning divisible goods. *Economic Theory*, 19(4):673–705, 2002.

Martin L Weitzman. Prices vs. quantities. *The review of economic studies*, 41(4):477–491, 1974.

Robert Wilson. Auctions of shares. *The Quarterly Journal of Economics*, pages 675–689, 1979.

Robert Wilson. Supply function equilibrium in a constrained transmission system. *Operations research*, 56(2):369–382, 2008.

H Witsenhausen. A minimax control problem for sampled linear systems. *IEEE Transactions on Automatic Control*, 13(1):5–21, 1968.

Hans S Witsenhausen. Minimax control of uncertain systems. In *IEEE Trans. Automat. Contr*. Citeseer, 1966.

APPENDIX A

Proof of Properties in Section (2.3)

We first present some preliminaries required for the proofs in Section (2.3.5). We start with the auction for elastic demand and then move to auctions with inelastic demand. Stated below, we present the GenCos' best response problems in the auction, Problem ($\hat{Q}_{1,n}^{elas.}$) and Problem ($\hat{Q}_{1,n}^{inelas.}$). We show that both the socially optimal problems, Problem (Socially Optimal Dispatch - Elastic) and Problem (Socially Optimal Dispatch - Inelastic), and the GenCos' best response problems, Problem ($\hat{Q}_{1,n}^{elas.}$) and Problem ($\hat{Q}_{1,n}^{inelas.}$), are strictly convex. Therefore, Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient conditions for optimality in all these problems. We present the KKT conditions.

A.1 Preliminaries

A set of messages $m^* = (m_1^*, m_2^*, ..., m_N^*)$ is a NE of the game induced by the proposed mechanism if for each producer $n \in N$, $m_n^* = (\hat{e}_n^*, p_n^*)$ is a best response to the others' messages denoted by m_{-n}^*.

First consider the auction with elastic demand. The pair $(\hat{e}_n^*, \hat{p}_n^*)$ is producer n's best response to other producers' messages, if it is a solution to his profit maximization problem

$$(\hat{e}_n^*, \hat{p}_n^*) = \arg \max_{\hat{e}_n^* \in \mathbb{R}, \hat{p}_n^* \in \mathbb{R}^+} \left\{ -C_n^e(\hat{e}_n) + r_n^{elas.} \right\} \qquad (\hat{Q}_{1,n}^{elas.})$$

$$s.t. \qquad \hat{p}_n \geq 0. \qquad (A.1)$$

We prove Problem $\hat{Q}_{1,n}^{elas.}$ is a strictly convex optimization. The domain of Problem $\hat{Q}_{1,n}^{elas.}$ is convex and compact. Furthermore, the inequality constraints are linear, therefore, they satisfy the constraint qualification conditions. The objective function is

$$-C_n^e(\hat{e}_n) + r_n = -C_n^e(\hat{e}_n) + \hat{p}_{n+1}\hat{e}_n - \hat{p}_n^{-0.5} {\zeta_n^{elas.}}^2 \qquad (A.2)$$

The term $-C_n^e(\hat{e}_n)$ is strictly concave by Assumption (A5). $\hat{p}_{n+1}\hat{e}_n$ is linear and concave. It remains to show the concavity of $-\hat{p}_n^{-0.5} {\zeta_n^{elas.}}^2$ which is equivalent to the convexity of $\hat{p}_n^{-0.5} {\zeta_n^{elas.}}^2$. The Hessian matrix of $\hat{p}_n^{-0.5} {\zeta_n^{elas.}}^2$ is

$$H = \begin{bmatrix} 0.75\hat{p}_n^{-2.5} {\zeta_n^{elas.}}^2 & -\hat{p}_n^{-1.5}(D(\hat{p}_{n+1}) - \sum_{n \in N} \hat{e}_n) \\ -\hat{p}_n^{-1.5}(D(\hat{p}_{n+1}) - \sum_{n \in N} \hat{e}_n) & 2\hat{p}_n^{-0.5} \end{bmatrix}. \qquad (A.3)$$

For convexity we need this Hessian to be positive semidefinite at any point $\hat{p}_n$ and $\hat{e}_n$ in the domain. Equivalently, the Hessian must have non-negative eigenvalues, that is, the following inequalities must be satisfied.

$$0.75\hat{p}_n^{-2.5}\zeta_n^{elas.2} + 2\hat{p}_n^{-0.5} \geq 0 \tag{A.4}$$

$$0.75\hat{p}_n^{-2.5}\zeta_n^{elas.2} \times 2\hat{p}_n^{-0.5} - \big(-\hat{p}_i^{-1.5}(D(\hat{p}_{n+1}) - \sum_{n\in N}\hat{e}_n)\big)^2 \geq 0 \tag{A.5}$$

Inequalities (A.4) and (A.5) are equivalent to

$$.75\hat{p}_n^{-2.5}\zeta_n^{elas.2} + 2\hat{p}_n^{-0.5} \geq 0 \tag{A.6}$$

$$0.5\hat{p}_n^{-3}\left(D(\hat{p}_{n+1}) - \sum_{n\in N}\hat{e}_n\right)^2 = 0.5\hat{p}_n^{-3}\zeta_n^{elas.2} \geq 0 \tag{A.7}$$

Inequalities (A.6) and (A.7) are satisfied in the domain of Problem $\hat{Q}_{1,n}^{elas.}$. Consequently, $\hat{p}_n^{-0.5}\zeta_n^{elas.2}$ is convex in $(\hat{p}_n, \hat{e}_n)$. Therefore, each individual producer's problem in the game induced by the mechanism is concave optimization.

Since both Problems (Socially Optimal Dispatch - Elastic) and ($\hat{Q}_{1,n}^{elas.}$) are strictly concave optimization problems, they have a unique solution, and the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality. The Lagrangian for problem (Socially Optimal Dispatch - Elastic) is

$$L_{SociallyOptimalDispatch-Elastic} = U(\sum_{n\in N} e_n) - \sum_{n\in N} C_n^e(e_n) \tag{A.8}$$

and the KKT conditions are, $\forall n \in N$,

$$\left.\frac{\partial U(\sum_{n\in N} e_n)}{\partial e_n}\right|_{e^*} - \left.\frac{\partial C_n^e}{\partial e_n}\right|_{e^*} = 0. \tag{A.9}$$

The Lagrangian for Problem ($\hat{Q}_{1,n}^{elas.}$) is

$$L_{\hat{Q}_{1,n}} = -C_n^e(\hat{e}_n) + r_n^{elas.} - \theta_n p_n, \tag{A.10}$$

and the corresponding KKT conditions are

$$-\left.\frac{\partial C_n^e}{\partial \hat{e}_n}\right|_{m^*} + \left.\frac{\partial r_n^{elas.}}{\partial \hat{e}_n}\right|_{m^*} = 0, \tag{A.11}$$

$$\frac{\partial r_n^{elas.}}{\partial \hat{p}_n}|_{m^*} - \hat{\theta}_n^* = 0, \tag{A.12}$$

$$\hat{\theta}_n^*\hat{p}_n^* = 0, \tag{A.13}$$

$$\hat{\theta}_n^* \geq 0. \tag{A.14}$$

Next we consider the auction for inelastic demand. Let $m^* = (\hat{e}_1^*, \hat{e}_2^*, ..., \hat{e}_N^*, \hat{p}_1^*, \hat{p}_2^*, ..., \hat{p}_N^*)$ be a NE of the game induced by the mechanism. Then, m^* is a solution to every producer's profit maximization problem, that is, $\forall n \in I$

$$(\hat{e}_n^*, p_n^*) = \arg\max_{\hat{e}_n, \hat{p}_n} -C_n^e(\hat{e}_n) + r_n^{inelas.} \qquad (\hat{Q}_{1,n}^{inelas.})$$

$$s.t \quad \hat{p}_n \geq 0. \qquad \text{(A.15)}$$

By an argument similar to the one used for the case of elastic demand, we can show that Problem ($\hat{Q}_{1,n}^{inelas.}$) is strictly convex optimization. Since both Problems (Socially Optimal Dispatch - Inelastic) and ($\hat{Q}_{1,n}^{inelas.}$) are strictly concave optimization, they have a unique solution and the KKT conditions are necessary and sufficient for optimization. The Lagrangian for Problem (Socially Optimal Dispatch - Inelastic) is

$$L_{SociallyOptimalDispatch-Inelastic} = -\sum_{i \in I} C_n^e e_n) + \kappa(\sum_{n \in N} e_n - \underline{D}) \qquad \text{(A.16)}$$

and KKT conditions are

$$\left.\frac{\partial C_n^e}{\partial e_n}\right|_{e^*} - \kappa^* = 0 \qquad \forall n \in N, \qquad \text{(A.17)}$$

$$\kappa^*(\sum_n e_n^* - \underline{D}) = 0, \qquad \text{(A.18)}$$

$$\kappa^* \geq 0. \qquad \text{(A.19)}$$

The Lagrangian for Problem ($\hat{Q}_{1,n}^{inelas.}$) is

$$L_{\hat{Q}_{1,n}^{inelas.}} = -C_n^e(\hat{e}_n) + r_n^{inelas.} + \hat{\theta}_n \hat{p}_n \qquad \text{(A.20)}$$

and the KKT conditions are, $\forall n \in N$,

$$-\left.\frac{\partial C_n^e}{\partial \hat{e}_n}\right|_{\mathbf{m}^*} + \left.\frac{\partial r_n^{inelas.}}{\partial \hat{e}_n}\right|_{\mathbf{m}^*} = 0, \qquad \text{(A.21)}$$

$$\frac{\partial r_n^{inelas.}}{\partial p_n}|_{\mathbf{m}^*} + \hat{\theta}_n^* = 0, \qquad \text{(A.22)}$$

$$\hat{\theta}_n^* \geq 0. \qquad \text{(A.23)}$$

A.2 Proof of Results in Section (2.3.5)

Proof of Lemma 1 First consider the auction for elastic demand. Assume $\zeta_n^{elas.*} \neq 0$ at NE for some $n \in N$. Then, the price at equilibrium, $\hat{p}_n^*$, must be strictly positive because if it is zero, the GenCo n's utility at equilibrium is equal to $-\infty$ and agent n can improve its utility by changing $\hat{p}_n$ to a positive value. From Eqs. (A.12) we have

$$\left.\frac{\partial u_n}{\partial \hat{p}_n}\right|_{m^*} = \left.\frac{\partial r_n^{elas.}}{\partial \hat{p}_n}\right|_{m^*} = -0.5\hat{p^*}_n^{-1.5} \left.\zeta_n^{elas.*2}\right|_{m^*} = 0. \qquad \text{(A.24)}$$

Since $\hat{p^*}_n^{-1.5} > 0$, we should have $\zeta_n^{elas.} = (D(\hat{p}_{n+1}) - \sum_{n\in N} \hat{e^*}_n)^2 = 0$ which contradicts the assumption of $\zeta_n^{elas.*} \neq 0$.

Next, consider the auction for inelastic demand. Since the equilibrium is non-trivial all prices can not be zero at the same time, therefore, either all prices are positive or there exists a zero price followed by a positive price.

Case 1: For all $n \in N$, $\hat{p}_n^* > 0$.

Here, $\hat{p}_n^* \geq 0$ is not binding. Therefore, at NE for all $n \in N$,

$$\left.\frac{\partial \hat{u}_n}{\partial \hat{p}_n}\right|_{m^*} = \left.\frac{\partial r_n^{inelas.}}{\partial \hat{p}_n}\right|_{m^*} = \\ -2(\hat{p}_n^* - \hat{p}_{n+1}^*) - 2\zeta^{inelas.*2} = 0. \tag{A.25}$$

Summing up Eq. (A.25) over all n we get:

$$\sum_{n\in N} \left.\frac{\partial r_n^{inelas.}}{\partial \hat{p}_n}\right|_{m^*} \\ = \sum_{n\in N} [-2(\hat{p}_n^* - \hat{p}_{n+1}^*) - 2\zeta^{inelas.*2}] \\ = -2N\zeta^{inelas.*2} = 0. \tag{A.26}$$

The last equality is achieved if and only if $\zeta_t^{inelas.*} = 0$, i.e. if and only if $\sum_{n\in N} \hat{e}_n^* = \underline{D}$.

Case 2: $\exists n \in N$ *s.t.* $\hat{p}_n^* = 0$ and $\hat{p}_{n+1} > 0$.

We prove that this case cannot exist. Since the constraint $\hat{p}_n \geq 0$ is binding for $\hat{p}_n^* = 0$, at NE we must have

$$\left.\frac{\partial \hat{u}_n}{\partial \hat{p}_n}\right|_{m^*} \leq 0. \tag{A.27}$$

Furthermore, using $\hat{p}_n^* = 0$ we obtain

$$\left.\frac{\partial \hat{u}_n}{\partial \hat{p}_n}\right|_{m^*} = 2[\hat{p}_{n+1}^* - \zeta^{inelas.*2}] \leq 0. \tag{A.28}$$

Since $\hat{p}_{n+1}^* > 0$, because of Eq. (A.25),

$$\left.\frac{\partial \hat{u}_{n+1}}{\partial \hat{p}_{n+1}}\right|_{m^*} = 2(\hat{p}_{n+2}^* - \hat{p}_{n+1}^*) - 2\zeta^{inelas.*2} = 0 \tag{A.29}$$

From (A.29) it follows that

$$\hat{p}_{n+2}^* = \hat{p}_{n+1}^* + \zeta^{inelas.*2} \geq \hat{p}_{n+1}^* > 0. \tag{A.30}$$

Since $\hat{p}_{n+2}^* > 0$, we can repeat the same argument as Eqs. (A.29) and (A.30) to calculate $\hat{p}_{n+3}^*$ and show it is positive. Continuing the same argument on the sequence of the prices

$\{\hat{p}^*_{n+3}, \hat{p}^*_{n+4}, ..., \hat{p}^*_n, \hat{p}^*_1, ..., \hat{p}^*_{n-1,t}, \hat{p}^*_n\}$, we obtain

$$\hat{p}^{*2}_n = \hat{p}^*_{n+1} + N\zeta^{inelas.*2} > 0. \tag{A.31}$$

This contradicts the assumption of $\hat{p}^*_n = 0$.

Proof of Theorem 1 First consider the auction for elastic demand. From Eq. (2.43) of Lemma 1, we have

$$D(\hat{p}^*_n) = \sum_{j \in N} \hat{e}^*_j \qquad \forall n \in N. \tag{A.32}$$

As a direct consequence of Eqs. (2.36) and (A.32),

$$\hat{p}^*_n = \hat{p}^* = U'(\sum_{j \in N} e^*_j) > 0 \qquad \forall n \in N. \tag{A.33}$$

The inequality in Eq. (A.33) follows from concavity of demand. Moreover, Eqs. (2.34), (2.43), (2.46) imply $r^*_n = \hat{p}^* \hat{e}^*_n$, and because $\zeta^{elas.*}_n = 0$,

$$\left.\frac{\partial r^{elas.}_n}{\partial \hat{e}_n}\right|_{m^*} = \hat{p}^* - 2\hat{p^*}^{0.5} \zeta^{elas.*}_n \left.\frac{\partial \zeta^{elas.}_n}{\partial \hat{e}_n}\right|_{m^*} = \hat{p}^*. \tag{A.34}$$

From Eq. (A.34), (A.11), we have

$$C'(e^*_n) = p^* \tag{A.35}$$

Next, consider the auction with inelastic demand. First note that from proof of Lemma 1, for any non-trivial NE, we have $p^*_n > 0, \forall n \in N$. Using Eq. (2.44) in Eq. (A.25) we obtain

$$\left.\frac{\partial r^{inelas.}_n}{\partial \hat{p}_n}\right|_{m^*} = -2(\hat{p}^*_n - \hat{p}^*_{n+1}) = 0. \tag{A.36}$$

Therefore,

$$\hat{p}^*_n = \hat{p}^*_{n+1} = \hat{p}^* \qquad \forall n \in N. \tag{A.37}$$

Next, from Eq. (2.45), we get $[\sum_{n \in N} \hat{e}^*_n - \underline{D}] = 0$, therefore,

$$\hat{p}^*[\sum_{n \in N} \hat{e}^*_n - \underline{D}] = 0 \quad \forall n \in N. \tag{A.38}$$

Furthermore, Eqs. (2.34), (2.44) and (A.37) imply

$$r^{inelas.*}_n = \hat{p}^* \hat{e}^*_n, \tag{A.39}$$

and

$$\left.\frac{\partial r^{inelas.}_n}{\partial \hat{e}_n}\right|_{m^*} = \hat{p}^* - 2\hat{p}^* \zeta^{inelas.*} \frac{\partial \zeta^{inelas.}}{\partial \hat{e}_n}|_{m^*} = \hat{p}^*. \tag{A.40}$$

Finally, using Eq. (A.40) in Eq. (A.21) of the KKT condition we obtain

$$\hat{p}^* = C_n^{e\prime}(\hat{e}_n^*). \tag{A.41}$$

Proof of Theorem 2 At the trivial equilibrium amount paid to the GenCos is zero. Therefore, the auctions are budget balanced. We discuss the case of non-trivial equilibriums below.

First consider the auction with elastic demand. At equilibrium, producer n receives $r_n^{elas.*} = p^* e_n^*$ and demand pays $\hat{p}^* \sum_{n\in N} \hat{e}_n^* = \sum_{n\in N} r_n^{elas.*}$ where from Eq. (A.33), $p^* = (U^{-1})'(\sum_{n\in N} e_n^*)$; therefore, the sum of all payments adds up to zero. For the auction with inelastic demand, the amount charged to producer, $\sum_{n\in N} r_n^{elas.*}$, can be collected from the inelastic demand without changing the behavior of the demand. Therefore, the auction is budget balanced.

Proof of Theorem 3 To show that the allocation $e_n^* = \hat{e}_n^*$, $n \in N$, corresponding to non-trivial equilibrium m^* is a solution of Problem (Socially Optimal Dispatch - Elastic) (correspondingly Problem (Socially Optimal Dispatch - Inelastic)), we specify the KKT parameters of Problem (Socially Optimal Dispatch - Elastic) (correspondingly Problem (Socially Optimal Dispatch - Inelastic)) based on the KKT parameters of Problem ($\hat{Q}_{1,n}^{elas.}$) (correspondingly Problem ($\hat{Q}_{1,n}^{inelas.}$)) as follows.

First consider the auction for elastic demand. Let m^* be a NE of the game induced by the proposed mechanism ($\hat{p}_n^* > 0, \quad \forall n \in N$ by Theorem 1). Then, Eqs. (A.11)-(A.13) along with Theorem 1 show that Eq. (A.9) is satisfied for all $n \in N$. Consequently, $(\hat{e}_1^*, \hat{e}_2^*, ..., \hat{e}_N^*)$ is a solution of Problem Socially Optimal Dispatch - Elastic.

Second, consider the auction for inelastic demand. Let m^* be a non-trivial NE of the game induced by the mechanism. Then, m^* is a solution to every GenCo's profit maximization problem. We set

$$e_n^* = \hat{e}_n^*, \quad \kappa^* = \hat{p}^* \quad \forall n \in N. \tag{A.42}$$

Then, Eqs. (A.21)-(A.23) along with Theorem 1 show that Eqs. (A.17)-(A.18) are satisfied.

Proof of Theorem 4 We show that each producer $n \in N$ weakly prefers the equilibrium outcome to the outside option $(e_n, r_n) = (0,0)$ resulting when producer n does not participate in the auction.

From Eq. (2.47) we obtain that GenCo n's utility at equilibrium is

$$-C_n^e(\hat{e}_n^*) + r_n^{elas.*} = -C_n^e(\hat{e}_n^*) + C_n^{e'}(\hat{e}_n^*) \times \hat{e}_n^*. \tag{A.43}$$

Furthermore, for the convex and increasing function $C_n^e(.)$ with $C_n^e(0) = 0$ we have

$$C_n^e(e_n) \le C_n^{e'}(e_n) \times e_n \qquad \text{for any} \quad e_n \ge 0. \tag{A.44}$$

Combining (A.43) and (A.44) we get,

$$u_n(m^*) = -C_n^e(\hat{e}_n^*) + C_n^{e'}(\hat{e}_n^*)\hat{e}_n^* \ge 0. \tag{A.45}$$

Proof of Lemma 2 We first prove the existence of a unique trivial equilbirum for the inelastic demand. Consider GenCo n and assume all other GenCos propose a price of zero; we prove GenCo n's best response is to bid zero price and zero quantity. Since $\hat{p}_{n+1}^* = 0$, GenCo n is paid no compensation for his production, i.e. $\hat{p}_{n+1}^* e_n = 0$. Other terms in GenCo n's utility function include penalty term and cost of production. The penalty terms are non-positive for any $\hat{p}_n > 0$ and zero for $\hat{p}_n = 0$. Therefore, producer n maximizes this utility by proposing $\hat{p}_n = 0$. Also, the cost of production is positive for all $\hat{e}_n \neq 0$ and zero at $\hat{e}_n = 0$. Therefore, GenCo n's best response is $e_n = 0$.

For elastic demand, the proof of Lemma (1) and Theorem (3) is valid at all equilibria[1]. Therefore, in the case where the solution of Problem (Socially Optimal Dispatch - Elastic) is such that $\exists n \in N : e_n^* > 0$ there does not exist any trivial equilibrium. This is because from Theorem (1), GenCos' price proposals are $\hat{p}^* = C_n^{e\prime}(e_n^*)$ which is greater than zero. On the other hand, if the solution of Problem (Socially Optimal Dispatch - Elastic) satisfies $e_n^* = 0, \forall n \in N$ then the price proposals are also zero and the game induced by the mechanism has a unique equilibrium which is trivial.

[1]Note that in contrast to auctions for elastic demand, for the case of inelastic demand we used the assumption of non-trivial equilibrium in the proof of Lemma (1).

APPENDIX B

Proof of Properties in Section (2.4)

We first present some preliminaries needed for the proofs of the results of Section (2.4). This preliminary information includes (i) the decomposition of Problem (Socially Optimal - Individual Const.) to optimization problems for each t and t^+, Problems $Q_2(1)$ and $Q_2(1^+)$, using dynamic programming principles, (ii) the best response problem for each GenCo at times 1 (capacity expansion) and 1^+ (generation spot market), Problems $\hat{Q}_{2,n}(1)$ and $\hat{Q}_{2,n}(1^+)$. We prove these problems are all strictly concave optimization. We present the KKT conditions that are necessary and sufficient for optimality.

B.1 Preliminaries for Proofs

Problem (Socially Optimal - Individual Const.)

For Problem (Socially Optimal - Individual Const.), we define the stage problems $Q_2(1)$ & $Q_2(1^+)$ respectively.

$Q_2(1^+)$

At time 1^+ optimal generations are solution to the following optimization problem,

$$V_{1^+}(h_{1^+}) = \max_{e_n, n\in N} \sum_{n\in N} U(\sum_{n\in N} e_n) - \sum_{n\in N} C_n^e(e_n) \tag{B.1}$$

$$s.t. \quad 0 \le e_n \le x_n \forall n \in N. \tag{B.2}$$

Note that $h_{1^+} = (x_0, \Delta x) = (x_{1,0}, x_{2,0}, \dots, x_{N,0}, \Delta x_1, \Delta x_2, \dots, \Delta x_N)$.

$Q_2(1^+)$ is a strictly concave optimization problem. The corresponding Lagrangian and KKT conditions are

$$\begin{aligned} L_{1^+}(h_{1^+}) = U(\sum_{n\in N} e_n) - \sum_{n\in N} C_n^e(e_n) \\ +\nu_n e_n + \mu_n(x_n - e_n). \end{aligned} \tag{B.3}$$

and $\forall n \in N$,

$$\frac{\partial U(\sum_{n\in N} e_n)}{\partial e_n}\Big|_{e^*} - \frac{\partial C_n^e(e_n)}{\partial e_n}\Big|_{e^*} + \nu_n^* - \mu_n^* = 0 \tag{B.4}$$

$$\nu_n^* e_n^* = 0 \tag{B.5}$$

$$\mu_n^*(x_n - e_n^*) = 0 \tag{B.6}$$

$$\nu_n^* \geq 0, \quad \mu_n^* \geq 0, \tag{B.7}$$

respectively.

$\underline{Q_2(1)}$

The capacity expansions at time 1 are the solution of the following optimization problem,

$$V_1(h_1) = \max_{\Delta x_n, n\in N} - \sum_{n\in N} C_n^x(\Delta x_n) + V_{1^+}(h_1, \Delta x) \tag{B.8}$$

$$s.t. \quad 0 \leq \Delta x_n \leq \overline{\Delta x}_n. \tag{B.9}$$

Note that $h_1 = x_0$.

We first prove $Q_2(1)$ is strictly concave optimization and then present the Lagrangian and KKT conditions. In order to prove strict concavity of $Q_2(1)$, we show that $V_{1^+}(h_1, \Delta x)$ is concave with respect to Δx for every h_1. We prove this in two steps. We first prove that $V_{1^+}(h_1, \Delta x)$ can be written as a function of $S_{1^+} = x_0 + \Delta x$. Next, we prove that $V_{1^+}(x_0 + \Delta x)$ is concave with respect to Δx.

First note that in the optimization problem specified by Eqs. (B.1)-(B.2), V_{1^+} depends on $x = x_0 + \Delta x$ in Equation (B.2). Therefore, $V_{1^+}(h_1, \Delta x)$ can be written as a function of $S_1 = x$. Next, we prove concavity of $V_{1^+}(x_0 + \Delta x)$ with respect to Δx. To do so, we first prove its concavity with respect to x. Consider x^1, x^2 and $x^3 = \lambda x^1 + (1-\lambda)x^2$ where $0 \leq \lambda \leq 1$ and their corresponding optimal production of electricity in Eq. (B.1), e^{1^*}, e^{2^*} and e^{3^*}. Consider $e^3 = \lambda e^{1^*} + (1-\lambda)e^{2^*}$. Since both e^{1^*} and e^{2^*} satisfy Eq. (B.2), $0 \leq e_n^3 \leq \lambda x_n^1 + (1-\lambda)x_n^2 = x_n^3$ $\forall n \in N$. This means e^3 is a feasible point in the domain of optimization of $V_{1^+}(x^3)$. In addition, since the objective function of function $V_{1^+}(x)$ is concave, then

$$\lambda V_{1^+}(x^1) + (1-\lambda)V_{1^+}(x^2) = \lambda[U(\sum_{n\in N} e_n^{1^*}) - \sum_{n\in N} C_n^e(e_n^{1^*})] + (1-\lambda)[U(\sum_{n\in N} e_n^{2^*}) - \sum_{n\in N} C_n^e(e_n^{2^*})] \tag{B.10}$$

$$\leq U(\sum_{n\in N} e_n^3) - \sum_{n\in N} C_n^e(e_n^3) \leq U(\sum_{n\in N} e_n^{3^*}) - \sum_{n\in N} C_n^e(e_n^{3^*}) = V_{1^+}(x^3) \tag{B.11}$$

This means V_{1^+} is concave. To prove concavity of $V_{1^+}(x_0 + \Delta x)$ with respect to Δx, define the function $B(\Delta x) = x_0 + \Delta x$. $B(\Delta x)$ is concave and $V_{1^+}(x)$ is non-decreasing (because by increasing x the domain of the optimization for V_{1^+} only increases, this implies that the value function is non-decreasing) and therefore, $V_{1^+}(B(\Delta x))$ is concave (See Rockafellar (2015), Theorem 5.1).

The corresponding Lagrangian and KKT conditions are

$$L(x_0) = -\sum_{n\in N} C_n^x(\Delta x_n) + V_{1^+}(x_0 + \Delta x) + \sum_{n\in N} \left(\alpha_n \Delta x_n + \beta_n(\overline{\Delta x}_n - \Delta x_n)\right) \tag{B.12}$$

and

$$-\frac{\partial C_n^x(\Delta x_n^*)}{\Delta x_n^*} + \frac{V_{1^+}^*(x_0 + \Delta x^*)}{\Delta x_n^*} + \alpha_n^* - \beta_n^* = 0, \tag{B.13}$$

$$\alpha_n^* \Delta x_n^* = 0, \tag{B.14}$$

$$\beta_n^*(\overline{\Delta x}_n - \Delta x_n^*) = 0, \tag{B.15}$$

$$\alpha_n^* \geq 0, \quad \beta_n^* \geq 0. \tag{B.16}$$

Best Response of the GenCos

For GenCo n, $n \in N$, we define the best response problems $\hat{Q}_{2,n}(1)$ and $\hat{Q}_{2,n}(1^+)$ at times 1 and 1^+ respectively. The history $\hat{h}_1$ at time 1 (resp. the history $\hat{h}_{1^+}$ at time 1^+) is an information state for problem $\hat{Q}_{2,n}(1)$ (respectively Problem $\hat{Q}_{2,n}(1^+)$). For each problem, we prove it is strictly concave optimization and present the corresponding Lagrangian and KKT conditions.

Let $\hat{m}_1^*$ and $\hat{m}_{1^+}^*$ be the GenCos' message at equilibrium.

$\hat{Q}_{2,n}(1^+)$

Consider GenCo $n \in N$ and denote the other GenCos' generation and price bid from their strategies at time 1^+ by $\hat{m}_{-n,1^+} = (\hat{e}_{-n}, \hat{p}_{-n})$. GenCo n's generation decision at the spot market of time 1^+ with history $\hat{h}_{1^+}$ is determined by the solution of the optimization problem

$$\hat{V}_{n,1^+}(\hat{h}_{1^+}) = \max_{\hat{e}_n, \hat{p}_n} -C_n^e(\hat{e}_n) + \hat{p}_{n+1}\hat{e}_n - \hat{p}_n^{-0.5}\zeta_n^{elas.\,2} \tag{B.17}$$

$$s.t. \quad 0 \leq \hat{e}_n \leq x_n \tag{B.18}$$

$$0 \leq \hat{p}_n. \tag{B.19}$$

The above optimization is strictly concave.

The corresponding Lagrangian and KKT conditions are

$$\hat{L}_{n,1^+}(\hat{h}_{1^+}) = -C_n^e(\hat{e}_n) + \hat{p}_{n+1}\hat{e}_n - \hat{p}_n^{-0.5}\zeta_n^{elas.\,2} + \hat{\nu}_n\hat{e}_n \tag{B.20}$$

$$+ \hat{\mu}_n(x_n - \hat{e}_n) + \hat{\theta}_n\hat{p}_n, \tag{B.21}$$

and

$$-\frac{\partial C_n^e}{\partial \hat{e}_n}|_{\hat{e}_n^*} + \hat{p}_{n+1}^* - \frac{\partial \hat{p}_n^{-0.5}\zeta_n^{elas.\,2}}{\partial \hat{e}_n}|_{(\hat{m}_{-n,1^+}, \hat{e}_n^*, \hat{p}_n^*)} + \hat{\nu}_n^* - \hat{\mu}_n^* = 0 \tag{B.22}$$

$$-\frac{\partial \hat{p}_n^{-0.5}\zeta_n^{elas.\,2}}{\partial \hat{p}_n}|_{(\hat{m}_{-n,1^+}, \hat{e}_n^*, \hat{p}_n^*)} + \hat{\theta}_n^* = 0, \tag{B.23}$$

$$\hat{\theta}_n^*\hat{p}_n^* = 0, \tag{B.24}$$

$$\hat{\mu}_n^*(x_n - \hat{e}_n^*) = 0, \tag{B.25}$$

$$\hat{\nu}_n^*\hat{e}_n^* = 0, \tag{B.26}$$

$$\hat{\nu}_n^* \geq 0, \quad \hat{\mu}_n^* \geq 0, \quad \hat{\theta}_n^* \geq 0, \tag{B.27}$$

respectively.

$\hat{Q}_{2,n}(1)$

Denote the other GenCos' expansion determined from their strategy at time 1 by $\hat{m}_{-n,1} = \Delta\hat{x}_{-n}$. GenCo n's best response for expansion decision at time 1, is determined by the solution of the optimization problem below:

$$\hat{V}_{n,1}(\hat{h}_1) = \max_{\Delta\hat{x}_n} -C^x_n(\Delta\hat{x}_n) + \hat{V}_{n,1^+}\big(\hat{h}_1, \Delta\hat{x}\big) \tag{B.28}$$

$$s.t. \quad 0 \leq \Delta\hat{x}_n \leq \overline{\Delta x}_n, \tag{B.29}$$

We prove that this is a strictly concave optimization problem. To do so, we prove that for every given $\hat{m}_{-n,1} = \Delta\hat{x}_{-n}$, $\hat{V}_{n,1^+}(\hat{h}_{1^+})$ is concave in $\Delta\hat{x}_n$. We show this in two steps. We first show that $\hat{V}_{n,1^+}(\hat{h}_{1^+})$ is a function of $\hat{x} = x_0 + \Delta\hat{x}$. To see this note that $\hat{V}_{n,1^+}(\hat{h}_{1^+})$ is a function of $\hat{h}_{1^+}$ because of the constraint described by Eq. (B.18). Next, for a given $\hat{x}_{-n}$, we prove $\hat{V}_{n,1^+}$ is concave in x_n and then we show it is concave in $\Delta\hat{x}_n$. Consider x^1, x^2 and $x^3 = \lambda x^1 + (1-\lambda)x^2$ where $0 \leq \lambda \leq 1$ and their corresponding optimal electricity and price bids in Eq. (B.17), $(\hat{e}^{1*}_n, p^{1*}_n)$, $(\hat{e}^{2*}_n, p^{2*}_n)$ and $(\hat{e}^{3*}_n, p^{3*}_n)$. Consider $\hat{e}^3_n = \lambda\hat{e}^{1*}_n + (1-\lambda)\hat{e}^{2*}_n$ and $p^3_n = \lambda p^{1*}_n + (1-\lambda)p^{2*}_n$. Since both $\hat{e}^{1*}_n$ and $\hat{e}^{1*}_n$ satisfy Eq. (B.18), $0 \leq \hat{e}^3_n \leq \lambda x^1_n + (1-\lambda)x^2_n = x^3_n \; \forall n \in N$. Also, $p^3_n \geq 0 \quad \forall n \in N$. This means $(\hat{e}^3_n, p^3_n)$ is a feasible point in the domain of optimization of $\hat{V}_{n,1^+}(x^3)$. In addition, since the objective function of function $\hat{V}_{n,1^+}(x)$ is concave, then

$$\begin{aligned}
&\lambda\hat{V}_{n,1^+}(x^1) + (1-\lambda)\hat{V}_{n,1^+}(x^2) = \lambda[-C^e_n(\hat{e}^{1*}_n) + \hat{p}^{1*}_{n+1}\hat{e}^{1*}_n - \hat{p}^{1*}_n{}^{-0.5}\zeta^{elas.1*}_n{}^2] \\
&+ (1-\lambda)[-C^e_n(\hat{e}^{2*}_n) + \hat{p}^{2*}_{n+1}\hat{e}^{2*}_n - \hat{p}^{2*}_n{}^{-0.5}\zeta^{elas.2*}_n{}^2] \qquad \text{(B.30)} \\
&\leq -C^e_n(\hat{e}^3_n) + \hat{p}^3_{n+1}\hat{e}^3_n - \hat{p}^3_n{}^{-0.5}\zeta^{elas.3}_n{}^2 \\
&\leq -C^e_n(\hat{e}^{3*}_n) + \hat{p}^{3*}_{n+1}\hat{e}^{3*}_n - \hat{p}^{3*}_n{}^{-0.5}\zeta^{elas.3*}_n{}^2 = \hat{V}_{n,1^+}(x_3) \qquad \text{(B.31)}
\end{aligned}$$

This means $\hat{V}_{1^+}$ is concave. To prove concavity of $\hat{V}_{1^+}(x_0 + \Delta x)$ with respect to Δx, define the function $B(\Delta x) = x_0 + \Delta x$. $B(\Delta x)$ is concave and $\hat{V}_{1^+}(x)$ is non-decreasing (because by increasing x the domain of the optimization for $\hat{V}_{1^+}$ only increases which means the value function is non-decreasing) and therefore, $\hat{V}_{1^+}(B(\Delta x))$ is concave (See Rockafellar (2015), Theorem 5.1).

The corresponding Lagrangian and KKT conditions are

$$\hat{L}_{n,1}(\hat{h}_1) = -C^x_n(\Delta\hat{x}_n) + \hat{V}_{1^+,n}\big(x_0 + \Delta\hat{x}\big) \tag{B.32}$$

$$+\hat{\alpha}_n\Delta\hat{x}_n + \hat{\beta}_n(\overline{\Delta x}_n - \Delta\hat{x}_n), \tag{B.33}$$

and

$$-\frac{\partial C^x_n(\Delta\hat{x}_n)}{\partial\Delta\hat{x}_n}\Big|_{\Delta\hat{x}^*_n} + \frac{\partial\hat{V}_{n,1^+}\big(x_0 + \Delta\hat{x}\big)}{\partial\Delta\hat{x}_n}\Big|_{(\hat{m}_{-n,1},\Delta\hat{x}^*_n)} + \hat{\alpha}^*_n - \hat{\beta}^*_n = 0, \tag{B.34}$$

$$\hat{\alpha}^*_n\Delta\hat{x}^*_n = 0, \tag{B.35}$$

$$\hat{\beta}^*_n(\overline{\Delta x}_n - \Delta\hat{x}^*_n) = 0, \tag{B.36}$$

$$\hat{\alpha}^*_n \geq 0, \quad \hat{\beta}^*_n \geq 0. \tag{B.37}$$

B.2 Proof of the Results in Section 2.4.3

Proof of Theorem (5) We first prove market clearance, budget balance and price efficiency. In the spot market. Problem $\hat{Q}_{2,n}(1^+)$ is similar to Problem $(\hat{Q}_{1,n}^{elas.})$ with additional individual constraint of Eq. (B.18); therefore, market clearance, budget balance and price efficiency of the spot market can be established in the similar way as in Problem $(\hat{Q}_{1,n}^{elas.})$ (Lemma (1) and Theorems (1) and (2)).

Next, we prove the social optimality of the equilibrium outcome. Assume $\hat{\sigma}_n^*$ and $\hat{\gamma}_n^*$ form a SPNE for the dynamic game induced by the mechanism proposed in Section (2.4.2). Consider $\hat{h}_1 = h_1$ and $\hat{h}_{1^+} = h_{1^+}$ and the following strategies in Problem (Socially Optimal - Individual Const.).

$$\sigma_n^*(S_1) = \Delta\hat{x}_{n,1^+}^*(S_{1^+}) \tag{B.38}$$

$$\gamma_n^*(S_{1^+}) = \hat{e}_n^*(S_{1^+}) \tag{B.39}$$

Then using the results of Theorem (3), Eq. (B.34)-(B.37), Eqs. (B.38)-(B.39) and Eqs. (B.22)-(B.27) the following parameters will satisfy the KKT conditions of $Q_2(1)$, Eq. (B.4)-(B.7), and the KKT conditions of $Q_2(1^+)$, Eq. (B.13)-(B.16), defined for all $n \in N$

$$\nu_n^* = \hat{\nu}_n^* \tag{B.40}$$

$$\mu_n^* = \hat{\mu}_n^* \tag{B.41}$$

$$\alpha_n^* = \hat{\alpha}_n^* \tag{B.42}$$

$$\beta_n^* = \hat{\beta}_n^* \tag{B.43}$$

This means the energy production and expansion corresponding to the SPNE are solutions of Problem (Socially Optimal - Individual Const.).

To prove that the mechanism is individually rational, we need to show that starting from each time 1 or 1^+, GenCo n's future utility due to the SPNE strategy of the game induced by the mechanism is greater than or equal to his non-continuation utility (which is 0).

First consider the expansion epoch 1. GenCo n's utility at equilibrium starting from time 1 is determined by the solution of $\hat{Q}_2(1)$ as

$$\hat{V}_n(h_1) = -C_n^e(\hat{e}_n^*) - C_n^x(\hat{x}_n^*) + \hat{p}^*\hat{e}_n^*. \tag{B.44}$$

To show this value is positive, we introduce a surrogate optimization problem which has the same solution and value function as $\hat{Q}_{2,n}(1)$. We prove the solution to this surrogate optimization problem is non-negative. Consider the following surrogate optimization for GenCo n where the prices p^* are fixed and equal to the outcome prices of the SPNE strategies.

$$\begin{aligned} \max_{\Delta\hat{x}_n', \hat{e}_n'} \quad & -C_n^e(\hat{e}_n') - C_n^x(\Delta\hat{x}_n') + \hat{p}^*\hat{e}_n' \\ s.t. \quad & 0 \le \Delta\hat{x}_n' \le \bar{x}_n \\ & 0 \le \hat{e}_n' \le x_{0,n} + \Delta\hat{x}_n \end{aligned} \tag{B.45}$$

Call this problem $\hat{Q}_2^{sur}(n)$. Problem $\hat{Q}_2^{sur}(n)$ is a convex optimization satisfying the constraint qualification conditions. First by using the KKT condition of Problem $\hat{Q}_2^{sur}(n)$, one can check that

GenCo n's SPNE strategies $\hat{\sigma}_n^*$ and $\hat{\gamma}_n^*$, are also a solution of $\hat{Q}_2^{sur}(n)$. Moreover, $\hat{V}_n$ is also equal to the optimal value of $\hat{Q}_2^{sur}(1)$. Then, note that $(0,0)$ is a feasible solution of problem $\hat{Q}_2^{sur}(n)$ and the value of $\hat{Q}_2^{sur}(n)$ at $(0,0)$ is equal to zero. Consequently, $\hat{V}_n$ is non-negative and GenCo n will continue participation in the mechanism from time 1.

Next, consider the generation at time 1^+. The utility at time 1^+ is determined by

$$\hat{V}_(n,1^+)(h_{1^+}) = -C_n^e(\hat{e}_n^*) + \hat{p}^*\hat{e}_n^*. \tag{B.46}$$

Similar to the proof of Theorem (4), one can show that $-C_n^e(\hat{e}_n^*) + \hat{p}^*\hat{e}_n^*$ is non-negative. Therefore, GenCo n will continue participation in the mechanism from time 1^+.

To prove uniqueness, note that Problem (Socially Optimal - Individual Const.) is strictly convex and therefore it has a unique solution. On the other hand, as shown earlier in this proof, at any SPNE of the game induced by the mechanism

$$\hat{e}_n^* = e_n^* \tag{B.47}$$

$$\Delta\hat{x}_n^* = \Delta x_n^* \tag{B.48}$$

and due to price efficiency

$$\hat{p}_n^* = U_t'(\sum_{n\in N} e_n^*). \tag{B.49}$$

Eqs. (B.47)-(B.49) uniquely determine the SPNE of the game induced by the mechanism.

Proof of Theorem 6: Theorem 5 shows that the SPNE of the game induced by the mechanism results in the same expansion and generation as the optimal solution of Problem (Socially Optimal - Individual Const.). Therefore to prove Theorem (6), we show that for Problem (Socially Optimal - Individual Const.) and its optimal strategy, if $\Delta x_n^* > 0$ GenCo n is saturated in the spot market, i.e.

$$e_n^* = x_n^*. \tag{B.50}$$

We prove this by contradiction. Assume $e_n^* < x_n^*$. We construct an alternative strategy which is feasible and can improve the social welfare. For the production strategy, use the same e_n^*. For the expansion strategy use $\Delta x_n^* - \varepsilon$ with small enough ε such that the constraints in Problem (Socially Optimal - Individual Const.) are not violated. This alternative strategy has lower expansion cost at time 1 compared to the original strategy but the cost for production and utility of demand remain the same. Therefore, the social welfare increases. This contradicts the fact that $(\gamma_n^*, \sigma_n^*, \quad n\in N)$ is the optimal solution of Problem (Socially Optimal - Individual Const.).

Next we show the price in the spot market covers the investment at its marginal cost. First, as proved in Theorem (5), by price efficiency the price in the spot market is equivalent to the marginal utility of the demand. Also, the expansions and the productions at the SPNE are equal to the expansions and productions in Problem (Socially Optimal - Individual Const.). Therefore, because of Eqs. (B.4) and (B.13), the price in the spot market is

$$p^* = C_n^{e\prime}(e_n^*) + C_n^{x\prime}(\Delta x_n^*) \text{ if } 0 < \Delta x_n < \overline{\Delta x}_n. \tag{B.51}$$

APPENDIX C

Proof of Properties in Section (2.5)

We first present some preliminary information needed for the proofs of the results of Section (2.6). This preliminary information includes (i) the decomposition of Problem (Socially Optimal - Longterm Uncertainty) to optimization problems for each t and t^+, Problems $Q_3(t)$ and $Q_3(t^+)$, using dynamic programming principles, (ii) the best response problem for each GenCo at times t (capacity expansion) and t^+ (generation spot market), Problems $\tilde{Q}_{3,n}(t)$ and $\tilde{Q}_{3,n}(t^+)$. We prove these problems are all strictly concave optimization. To do so, we prove Lemma (3). We present the KKT conditions that are necessary and sufficient for optimality.

C.1 Preliminaries

For Problem (Socially Optimal - Longterm Uncertainty) we define the stage problems $Q_3(t)$ & $Q_3(t^+)$ respectively.

Let

$$V_{T+1}(h_{T+1}) = \eta(w_T) \sum_{n\in N} x_{n,T+1}. \tag{C.1}$$

Problem $Q_3(t^+)$

At time t^+ optimal generations are according to the following optimization problem,

$$V_{t^+}(h_{t^+}, w_{1:t}) = \max_{e_{n,t}, n\in N} U_t(\sum_{n\in N} e_{n,t}, w_t) - \sum_{n\in N} C^e_{n,t}(e_{n,t}, w_t) + \beta V_{t+1}(h_{t^+}, e_t) \tag{C.2}$$

$$s.t. \quad 0 \le e_{n,t} \le x_{n,t^+} \qquad \forall n \in N. \tag{C.3}$$

Problem $Q_3(t)$

The expansions at time t are according to the following optimization problem which we refer to by $Q_3(t)$ (note that $(h_t, \Delta x_t, w_t) = h_{t^+}$).

$$V_t(h_t) = \max_{\Delta x_{n,t}, n\in N} -\sum_n C^x_{n,t}(\Delta x_{n,t}, w_{t-1}) + E_{\mathbf{w}_t}\left[V_{t^+}(h_t, \Delta x_t, \mathbf{w}_t)|w_{1:t-1}\right] \tag{C.4}$$

$$s.t. \quad 0 \le \Delta x_{n,t} \le \overline{\Delta x}_{n,t} \tag{C.5}$$

$$\sum_{n\in N} \Delta x_{n,t} \ge \underline{\Delta x}_t \tag{C.6}$$

GenCos' Best Response

For GenCo n, $n \in N$, we define best response problems $\tilde{Q}_{3,n}(t)$ and $\tilde{Q}_{3,n}(t^+)$ at times t and t^+ respectively. The history $\tilde{h}_t$ at time t (resp. the history $\tilde{h}_{t^+}$ at time t^+) is an information state for problem $\tilde{Q}_{3,n}(t)$ (respectively problem $\tilde{Q}_{3,n}(t^+)$).

Let

$$\tilde{V}_{n,T+1}(\tilde{h}_{T+1}) = \eta \tilde{x}_{n,T+1}. \tag{C.7}$$

$\tilde{Q}_{3,n}(t^+)$

At the spot market of time t^+ with history $\tilde{h}_{t^+}$, GenCo's best response to other GenCos' production and price proposal $\tilde{m}_{-n,t^+} = (\hat{e}_{-n,t}, \hat{p}_{-n,t})$ determined by their strategies $\hat{\gamma}_{-n,t}$ is

$$\tilde{V}_{n,t^+}(\tilde{h}_{t^+}) = \max_{\hat{e}_{n,t},\hat{p}_{n,t}} -C^e_{n,t}(\hat{e}_{n,t}, w_t) + \hat{p}_{n+1,t}\hat{e}_{n,t} - \hat{p}_{n,t}^{-0.5}\zeta^{elas.}{}^2_{n,t} + \beta \tilde{V}_{N,t+1}(\tilde{h}_t, \hat{e}_t, \hat{p}_t) \tag{C.8}$$

$$s.t. \quad 0 \leq \hat{e}_{n,t} \leq x_{n,t} \tag{C.9}$$

$$0 \leq \hat{p}_{n,t} \tag{C.10}$$

$\tilde{Q}_{3,n}(t)$

At expansion epoch t with history $\tilde{h}_t$, GenCo n's best response to the other GenCos' expansion and price proposal $\tilde{m}_{-n,t} = (\Delta\tilde{x}_{-n,t}, \tilde{p}_{-n,t})$ determined by their strategies $\tilde{\sigma}_{-n,t}$ is determined by

$$\tilde{V}_{n,t}(\tilde{h}_t) = \max_{\Delta\tilde{x}_{n,t},\tilde{p}_{n,t}} -C^x_{n,t}(\Delta\tilde{x}_{n,t}, w_{t-1}) + \tilde{p}^*_{n+1,t}\Delta\tilde{x}_{n,t} - (\tilde{p}_{n,t} - \tilde{p}^*_{n+1,t})^2 - 2\tilde{p}_{n,t}\zeta^{inelas.}{}^2_t$$

$$+ \beta E_{\mathbf{w}_t}\left[\tilde{V}_{t^+,n}\left(\tilde{h}_t, \Delta\tilde{x}_t, \tilde{p}_t, \mathbf{w}_t\right)\middle| w_{1:t-1}\right] \tag{C.11}$$

$$0 \leq \Delta\tilde{x}_{n,t} \leq \overline{\Delta x}_{n,t} \tag{C.12}$$

$$\tilde{p}_{n,t} \geq 0 \tag{C.13}$$

To show Problems $\tilde{Q}_{3,n}(t)$ and $\tilde{Q}_{3,n}(t^+)$ are strictly concave optimization and develop the Lagrangian and KKT conditions, we first prove Lemma (3).

Proof of Lemma 3 We show that $S_t := (x_{t-1}, w_{1:t-1})$ (respectively $S_{t^+} := (x_t, w_{1:t})$) is an information state for problem $Q_3(t)$ and $\tilde{Q}_{3,n}(t)$ (respectively $Q_3(t^+)$ and $\tilde{Q}_{3,n}(t^+)$). We first prove the results for Problem (Socially Optimal - Longterm Uncertainty) and then for the SPNE of the game induced by the mechanism. We proceed by induction.

Problem (Socially Optimal - Longterm Uncertainty)

We proceed in three steps. We first show that the value function at time $T+1$, V_{T+1} is a function of S_{T+1}. Then we prove that if for $t+1 \in T$, V_{t+1} is function of S_{t+1} then V_{t^+} is also a function of S_{t^+}. Finally, we show that if for $t \in T$, V_{t^+} is a function of S_{t^+} then V_t is a function of S_t.

First, note that at time $T+1$, V_{T+1} is a function of S_{T+1} (see Eq. C.1). Assume that $V_{\tau+1}$, $\tau = t+1, t+2, ..., T$, is a function of $S_{\tau+1}$. Then, at t^+ the solution of Eq. (C.2) is a function of $(x_t, w_{1:t}) = S_{t^+}$, therefore, S_{t^+} is an information state for Problem $Q_3(t^+)$ and V_{t^+} is a function of S_{t^+}. Consequently, the solution of Eq. (C.4) is a funciton of $(x_{t-1}, w_{1:t-1}) = S_t$, thus S_t is an information state for Problem $Q_3(t)$ and V_t is a funciton of S_t.

GenCos' best response problem

We prove in three steps. We first show that the value function at time $T+1$, $\tilde{V}_{n,T+1}$ is a function of S_{T+1}. Then we prove that if for $t+1 \in T$, $\tilde{V}_{n,t+1}$ is function of S_{t+1} for all $n \in N$ then $\tilde{V}_{n,t^+}$ is also a function of S_{t^+}. Finally, we show that if for $t \in T$, $\tilde{V}_{n,t^+}$ is a function of S_{t^+} for all $n \in N$, then $\tilde{V}_{n,t}$ is a function of S_t.

First, at time $T+1$, by Eq. (C.7), $\tilde{V}_{n,T+1}(\tilde{h}_{T+1})$ is only a function of S_{T+1} for all $n \in N$.

Second, at time t^+, $t \in T$, if $\tilde{V}_{n,t+1}$ is a function of S_{t+1}, it is independent of the strategies at time t^+ (because these strategies do not change S_{t+1}). On the other hand, the best response problem and its constraints at time t^+ is a function of S_t. Therefore, equilibrium strategies are only a function of S_t[1].

Third, we consider the expansion at time $t \in T$ and assume $\tilde{V}^*_{n,t^+}$ is a function of S_{t^+}; S_{t^+} is determined by S_t plus the expansion decision at time t (i.e. independent of the rest of the history). Therefore the best response problem at time t and its constraints can be written as a function of S_t. As a result, the equilibrium strategies are a function of S_t [2].

Using the above Lemma, we can prove strict concavity of the stage optimization problems and develop the KKT conditions.

Dynamic Programming and KKT Conditions for Problem
(Socially Optimal - Longterm Uncertainty)
First define

$$V_{T+1}(S_{T+1}) = \eta(w_T) \sum_{n \in N} x_{n,T+1}. \tag{C.14}$$

and note that it is concave.

Problem $Q_3(t^+)$
At time t^+ optimal generations are according to the following optimization problem,

$$V_{t^+}(x_{t^+}, w_{1:t}) = \max_{e_{n,t}, n \in N} U_t(\sum_{n \in N} e_{n,t}, w_t) - \sum_{n \in N} C^e_{n,t}(e_{n,t}, w_t) + \beta V_{t+1}(x_{t^+}, w_{1:t}) \tag{C.15}$$

$$s.t. \quad 0 \le e_{n,t} \le x_{n,t^+} \qquad \forall n \in N. \tag{C.16}$$

Assuming $V_{t+1}(x_{t+1}, w_{1:t})$ is concave in x_{t+1} (by induction), we prove V_{t^+} is a strictly concave optimization problem and furthermore, V_t is strictly concave with respect to x_t for every $w_{1:t} \in W_{1:t}$. For a give $w_{1:t}$, $C^e_{n,t}$ is strictly convex, U_t is strictly concave and by assumption V_{t+1} is concave. Since the domain is also concave, the problem is strictly concave optimization. The corresponding Lagrangian and KKT conditions are

$$L_{t^+}(S_{t^+}) = U_t(\sum_{n \in N} e_{n,t}, w_t) - \sum_{n \in N} C^e_{n,t}(e_{n,t}, w_t) + \beta V_{t+1}(S_{t^+}) \tag{C.17}$$

$$+ \sum_{n \in N} \nu_{n,t} e_{n,t} + \sum_{n \in N} \mu_{n,t}(x_{n,t^+} - e_{n,t}), \tag{C.18}$$

[1] In problems with multiple equilibrium, the equilibrium selection can be history dependent. We will show later that the spot market at time t has a unique equilibrium and therefore, there does not exist an equilibrium selection problem.

[2] We will show later that there are two equilibria in each capacity market which are both function of S_t. In general with multiple equilibria, equilibrium selection can be history dependent. Therefroe, here we consider Pareto efficient equilibrium. Since one of the two equilibria Pareto dominates the other, in this case, equilibrium selection is not a history dependent decision.

and

$$
\begin{aligned}
\frac{\partial U_t(\sum_{n\in N} e_{n,t}, w_t)}{\partial e_{n,t}}|_{e_t^*} - \frac{\partial C^e_{n,t}(e_{n,t}, w_t)}{\partial e_{n,t}}|_{e_t^*}& \\
+\beta \frac{\partial V_{t+1}(S_{t^+})}{\partial e_{n,t}}|_{e^*} + \nu^*_{n,t} - \mu^*_{n,t} = 0,& \quad \text{(C.19)}\\
\nu^*_{n,t} e^*_{n,t} = 0,& \quad \text{(C.20)}\\
\mu^*_{n,t}(x^*_{n,t^+} - e^*_{n,t}) = 0,& \quad \text{(C.21)}\\
\nu^*_{n,t} \geq 0, \quad \mu^*_{n,t} \geq 0,& \quad \text{(C.22)}
\end{aligned}
$$

respectively.

Next, we prove that $V_{t^+}(S_{t^+})$ is concave in x_{t^+}. Consider x_t^1, x_t^2, $x_t^3 = \lambda x_t^1 + (1-\lambda)x_t^2$, where $0 \leq \lambda \leq 1$ and their corresponding optimal production of electricity in Eq. (C.15), $e_t^{1^*}$, $e_t^{2^*}$ and $e_t^{3^*}$. Consider $e_t^3 = \lambda e_t^{1^*} + (1-\lambda)e_t^{2^*}$. Since both $e_t^{1^*}$ and $e_t^{2^*}$ satisfy Eq. (C.16), $0 \leq e^3_{n,t} \leq \lambda x_n^1 + (1-\lambda)x_n^2 = x_n^3\ \forall n \in N$. This means e^3 is a feasible point in the domain of optimization of V_{t^+} for x^3. In addition, since the objective function of function V_{t^+} is concave, we have

$$
\begin{aligned}
&\lambda V_{t^+}(x^1, w_{1:t}) + (1-\lambda)V_{t^+}(x^2, w_{1:t}) = \lambda[U_t(\sum_{n\in N} e^{1*}_{n,t}, w_t) - \sum_{n\in N} C^e_{n,t}(e^{1*}_{n,t}, w_t) + \beta V_{t+1}(x^1_{t^+}, w_{1:t})]\\
&+(1-\lambda)[U_t(\sum_{n\in N} e^{2*}_{n,t}, w_t) - \sum_{n\in N} C^e_{n,t}(e^{2*}_{n,t}, w_t) + \beta V_{t+1}(x^1_{t^+}, w_{1:t})]\\
&\leq U_t(\sum_{n\in N} e^{3}_{n,t}, w_t) - \sum_{n\in N} C^e_{n,t}(e^{3}_{n,t}, w_t) + \beta V_{t+1}(x^1_{t^+}, w_{1:t}) =\\
&\leq U_t(\sum_{n\in N} e^{3*}_{n,t}, w_t) - \sum_{n\in N} C^e_{n,t}(e^{3*}_{n,t}, w_t) + \beta V_{t+1}(x^1_{t^+}, w_{1:t}) = V_{t^+}(x^3, w_{1:t}) \qquad \text{(C.23)}
\end{aligned}
$$

This means V_{n,t^+} is concave in x_t.

Problem $Q_3(t)$

The expansions at time t are according to the following optimization problem which we refer to by $Q_3(t)$ (note that $(x_t + \Delta x_t, w_{1:t}) = S_{t^+}$).

$$
\begin{aligned}
V_t(S_t) = &\max_{\Delta x_{n,t}, n\in N} -\sum_n C^x_{n,t}(\Delta x_{n,t}, w_{t-1}) + E_{\mathbf{w}_t}\left[V_{t^+}(x_t + \Delta x_t, \mathbf{w}_{1:t})|w_{1:t-1}\right] \qquad \text{(C.24)}\\
s.t. \quad & 0 \leq \Delta x_{n,t} \leq \overline{\Delta x}_{n,t} \qquad \text{(C.25)}\\
& \sum_{n\in N} \Delta x_{n,t} \geq \underline{\Delta x_t} \qquad \text{(C.26)}
\end{aligned}
$$

Assuming that $V_{t^+}(x_t + \Delta x_t, \mathbf{w}_{1:t})$ is concave with respect to x_t (by induction), we prove $Q_3(t)$ is a strictly concave optimization problem and furthermore, $V_t(S_t)$ is concave in x_t. First, to prove concavity of $V_t(S_t)$ with respect to Δx_t, define the function $B(\Delta x_t) = x_t + \Delta x_t$. $B(\Delta x_t)$ is concave and $V_t(S_t)$ is concave and non-decreasing in x_t, therefore, $V_{t^+}(B(\Delta x_t))$ is concave in Δx_t (See Rockafellar (2015), Theorem 5.1) and $E_{\mathbf{w}_t}\left[V_{t^+}(x_{t^+} + \Delta x_t, \mathbf{w}_{1:t})|w_{1:t-1}\right]$ is a weighted sum of concave functions, hence it is concave in Δx_t. Considering this, it is straightforward to see $V_t(S_t)$ is concave optimization with convex domain. The Lagrangian and the KKT conditions are

$$L_t(S_t) = -\sum_n C^x_{n,t}(\Delta x_{n,t}, w_{t-1}) + \beta E_{\mathbf{w}_t}\left[V_{t^+}(x_t + \Delta x_t, \mathbf{w}_{1:t})\big| w_{1:t-1}\right] + \sum_{n\in N}\left(\alpha_{n,t}\Delta x_{n,t} + \beta_{n,t}(\overline{\Delta x}_{n,t} - \Delta x_{n,t})\right) + \kappa_t(\sum_{n\in N}\Delta x_{n,t} - \underline{\Delta x}_t) \tag{C.27}$$

and

$$-\frac{\partial C^x_{n,t}(\Delta x_{n,t}, w_{t-1})}{\partial \Delta x_{n,t}}|_{\Delta x_t^*} + \beta E_{\mathbf{w}_t}\left[\frac{\partial V_{t^+}(x_t+\Delta x_t, \mathbf{w}_{1:t})}{\partial \Delta x_{n,t}}|_{\Delta x_t^*}\big| w_{1:t-1}\right] + \alpha^*_{n,t} - \beta^*_{n,t} + \kappa^*_t = 0 \tag{C.28}$$

$$\alpha^*_{n,t}\Delta x^*_{n,t} = 0 \tag{C.29}$$

$$\beta^*_{n,t}(\overline{\Delta x}_{n,t} - \Delta x^*_{n,t}) = 0 \tag{C.30}$$

$$\kappa^*_t(\sum_{n\in N} x^*_{n,t} - \underline{\Delta x}_t) = 0 \tag{C.31}$$

$$\alpha^*_{n,t} \geq 0, \quad \beta^*_{n,t} \geq 0, \kappa^*_t \geq 0 \tag{C.32}$$

respectively.

To prove $V_t(S_t)$ is concave in x_t for a given $w_{1:t-1} \in W_{1:t-1}$, we use the first and the second order conditions. First, by envelope theorem, $\frac{\partial V_t(S_t)}{\partial x_{n,t}} = \beta E_{w_t}\left[\frac{\partial V_{t^+}(x_t+\Delta x_t, \mathbf{w}_{1:t})}{\partial x_{n,t}}|_{\Delta x_t^*}\big| w_{1:t-1}\right]$ which is non-negative by concavity of $V_{t^+}(x_t+\Delta x_t, \mathbf{w}_{1:t})$ with respect to $x_t + \Delta x_t$ (first order condition). Second,

$$Hessian(V_t(S_t)) = \beta E_{w_t}\left[Hessian(V_{t^+}(x_t + \Delta x_t, \mathbf{w}_{1:t}))|_{\Delta x_t^*}\big| w_{1:t-1}\right] \tag{C.33}$$

which is negative semidefinite by concavity of $V_{t^+}(x_t+\Delta x_t, \mathbf{w}_{1:t})$ (second order condition). First and second order conditions prove concavity of $V_t(S_t)$.

GenCos' Best Response

For GenCo n, $n \in N$ we present the best response problems $\tilde{Q}_{3,n}(t)$ and $\tilde{Q}_{3,n}(t^+)$ at times t and t^+ respectively as a function of S_t and S_{t^+}. Let

$$\tilde{V}_{n,T+1}(S_{T+1}) = \eta \tilde{x}_{n,T+1}. \tag{C.34}$$

$\tilde{V}_{n,T+1}(S_{T+1})$ is concave in $\tilde{x}_{n,T+1}$.

$\tilde{Q}_{3,n}(t^+)$

At the spot market of time t^+ with S_{t^+}, GenCo's best response to other GenCos' production and price proposal $\tilde{m}_{-n,t^+} = (\hat{e}_{-n,t}, \hat{p}_{-n,t})$ determined by their strategies $\hat{\gamma}_{-n,t}$ is

$$\tilde{V}_{n,t^+}(S_{t^+}) = \max_{\hat{e}_{n,t}, \hat{p}_{n,t}} -C^e_{n,t}(\hat{e}_{n,t}, w_t) + \hat{p}_{n+1,t}\hat{e}_{n,t} - \hat{p}_{n,t}^{-0.5} {\zeta^{elas.}}^2_{n,t} + \beta \tilde{V}_{n,t+1}(S_t) \tag{C.35}$$

$$s.t. \quad 0 \leq \hat{e}_{n,t} \leq x_{n,t} \tag{C.36}$$

$$0 \leq \hat{p}_{n,t} \tag{C.37}$$

S_t is independent of $\hat{e}_{n,t}, \hat{p}_{n,t}$ and therefore, $\beta \tilde{V}_{n,t+1}(S_t)$ is independent of $\hat{e}_{n,t}, \hat{p}_{n,t}$. Therefore, it is straightforward to see $\tilde{V}_{n,t}$ is concave optimization with convex domain. The Lagrangian and the KKT condition are

$$\tilde{L}_{n,t^+}(S_{t^+}) = -C^e_{n,t}(\hat{e}_{n,t}, \mathbf{w}_t) + \hat{p}_{n+1,t}\hat{e}_{n,t} - \hat{p}_{n,t}^{-0.5} {\zeta^{elas.}}^2_{n,t} + \beta \tilde{V}_{n,t+1}(S_{t^+}) + \hat{\nu}_{n,t}\hat{e}_{n,t} \tag{C.38}$$

$$+ \hat{\mu}_{n,t}(x_{n,t} - \hat{e}_{n,t}) + \hat{\theta}_{n,t}\hat{p}_{n,t}, \tag{C.39}$$

and

$$-\frac{\partial C^e_{n,t}}{\partial \hat{e}_{n,t}}\Big|_{\hat{e}^*_{n,t}} + \hat{p}^*_{n+1,t} - \frac{\partial \hat{p}_{n,t}^{-0.5} {\zeta^{elas.}}^2_{n,t}}{\partial \hat{e}_{n,t}}\Big|_{(\tilde{m}_{-n,t^+}, \tilde{e}_{n,t}, \tilde{p}_{n,t})}$$

$$+ \beta \frac{\partial \tilde{V}_{n,t+1}(S_t)}{\partial \hat{e}_{n,t}}\Big|_{(\tilde{m}_{-n,t^+}, \tilde{e}_{n,t}, \tilde{p}_{n,t})} + \hat{\nu}^*_{n,t} - \hat{\mu}^*_{n,t} = 0, \tag{C.40}$$

$$-\frac{\partial \hat{p}_{n,t}^{-0.5} {\zeta^{elas.}}^2_{n,t}}{\partial \hat{p}_{n,t}}\Big|_{(\tilde{m}_{-n,t^+}, \tilde{e}^*_{n,t}, \tilde{p}^*_{n,t})} + \beta \frac{\partial \tilde{V}_{n,t+1}(S_{t+1})}{\partial \hat{p}_{n,t}}\Big|_{(\tilde{m}_{-n,t^+}, \tilde{e}^*_{n,t}, \tilde{p}^*_{n,t})} + \hat{\theta}^*_{n,t} = 0, \tag{C.41}$$

$$\hat{\theta}^*_{n,t}\hat{p}^*_{n,t} = 0, \tag{C.42}$$

$$\hat{\mu}^*_{n,t}(x_{n,t} - \hat{e}^*_{n,t}) = 0, \tag{C.43}$$

$$\hat{\nu}^*_{n,t}\hat{e}^*_{n,t} = 0, \tag{C.44}$$

$$\hat{\nu}^*_{n,t} \geq 0, \quad \hat{\mu}^*_{n,t} \geq 0, \quad \hat{\theta}^*_{n,t} \geq 0, \tag{C.45}$$

respectively.

To prove that for every given $w_{1:t}$, $\tilde{V}_{n,t^+}(S_{t^+})$ is concave with respect to x_{n,t^+}, assume $\tilde{V}_{n,t+1}$ is concave in x_{t+1} for every $w_{1:t} \in W_{1:t}$ (by induction). Then by employing the arguments used to prove concavity of $\hat{V}_{n,1^+}$ in Eqs. (B.28)-(B.29), we can establish the concavity of $\tilde{V}_{n,t+1}$ with respect to x_{n,t^+}.

$\tilde{Q}_{3,n}(t)$

At expansion epoch t with S_t, GenCo n's best response to the other GenCos' expansion and price proposal $\tilde{m}_{-n,t} = (\Delta\tilde{x}_{-n,t}, \tilde{p}_{-n,t})$ determined by their strategies $\tilde{\sigma}_{-n,t}$ is determined by

$$\tilde{V}_{n,t}(S_t) = \max_{\Delta\tilde{x}_{n,t}} -C^x_{n,t}(\Delta\tilde{x}_{n,t}, w_{t-1}) + \tilde{p}^*_{n+1,t}\Delta\tilde{x}_{n,t} - (\tilde{p}_{n,t} - \tilde{p}^*_{n+1,t})^2 - 2\tilde{p}_{n,t} {\zeta^{inelas.}}^2_t$$

$$+ \beta E_{\mathbf{w}_t}\left[\tilde{V}_{t^+,n}\big(x_t + \Delta\tilde{x}_t, \mathbf{w}_{1:t}\big)\big|w_{1:t-1}\right] \tag{C.46}$$

$$0 \leq \Delta\tilde{x}_{n,t} \leq \overline{\Delta x}_{n,t} \tag{C.47}$$

$$\tilde{p}_{n,t} \geq 0. \tag{C.48}$$

We prove that the above optimization is a strictly concave optimization and furthermore, $\tilde{V}_{n,t}(S_t)$ is concave in x_t for every $w_{1:t} \in W_{1:t}$. To do so, assume (by induction) that $\tilde{V}_{n,t^+}(x_t + \Delta\tilde{x}_t, w_{1:t})$ is concave and increasing in $x_t + \Delta\tilde{x}_t$. Define function $B(\Delta\tilde{x}_t) = x_t + \Delta\tilde{x}_t$. Since $B(\Delta\tilde{x})$ is affine, this means $\tilde{V}_{n,t^+}(x_t + \Delta\tilde{x}_t, w_{1:t})$ is concave in $\Delta\tilde{x}$ (See Rockafellar (2015), Theorem 5.1).

Next, $E_{\mathbf{w}_{1:t}}\left[V_{n,t^+}(x_{t^+} + \Delta x_t, \mathbf{w}_t)|w_{1:t-1}\right]$ is a weighted sum of concave functions hence concave. Considering this, it is straightforward to see $V_{n,t}(S_t)$ is a strictly concave optimization with convex domain. The Lagrangian and the KKT conditions are

$$\begin{aligned}\tilde{L}_{n,t}(S_t) = -C^x_{n,t}(\Delta\tilde{x_{n,t}}, w_{t-1}) + \tilde{p}^*_{n+1,t}\Delta\tilde{x}_{n,t} - (\tilde{p}_{n,t} - \tilde{p}^*_{n+1,t})^2 \\ -2\tilde{p}_{n,t}\zeta^{inelas.2}_t + \beta E_{\mathbf{w}_t}\left[\tilde{V}_{n,t^+}\big(x_t + \Delta\tilde{x}_t, \mathbf{w}_{1:t}\big)\big|w_{1:t-1}\right] + \tilde{\alpha}_{n,t}\Delta\tilde{x}_{n,t} + \\ \tilde{\beta}_{n,t}(\overline{\Delta x}_{n,t} - \Delta\tilde{x}_{n,t}) + \tilde{\theta}_{n,t}\tilde{p}_{n,t},\end{aligned} \tag{C.49}$$

and

$$\begin{aligned}-\frac{\partial C^x_{n,t}(\Delta\tilde{x}_{n,t}, w_{t-1})}{\partial\Delta\tilde{x}_{n,t}}|_{\Delta\tilde{x}^*_{n,t}} + \frac{\partial r^{inelas.}_{n,t}}{\partial\Delta\tilde{x}_{n,t}}|_{(\tilde{m}_{-n,t},\Delta\tilde{x}^*_{n,t},\tilde{p}^*_{n,t})} \\ +\beta E_{\mathbf{w}_{1:t}}\left[\frac{\partial\tilde{V}_{n,t^+}\big(x_{t^+} + \Delta\tilde{x}_t, \mathbf{w}_t\big)}{\partial\Delta\tilde{x}_{n,t}}|_{(\tilde{m}_{-n,t},\Delta\tilde{x}^*_{n,t},\tilde{p}^*_{n,t})}\big|w_{1:t-1}\right], \\ +\tilde{\alpha}^*_{n,t} - \tilde{\beta}^*_{n,t} = 0,\end{aligned} \tag{C.50}$$

$$\frac{\partial r^{inelas.}_{n,t}}{\partial\tilde{p}_{n,t}}|_{(\tilde{m}_{-n,t},\Delta\tilde{x}^*_{n,t},\tilde{e}^*_{n,t})} + \beta E_{w_t}\left[\frac{\partial\tilde{V}_{n,t^+}\big(x_t + \Delta\tilde{x}_t, \mathbf{w}_{1:t}\big)}{\partial\tilde{p}_{n,t}}|_{(\tilde{m}_{-n,t},\Delta\tilde{x}^*_{n,t},\tilde{p}^*_{n,t})}\big|w_{1:t-1}\right], \tag{C.51}$$

$$+\tilde{\alpha}^*_{n,t}\Delta\tilde{x}^*_{n,t} = 0, \tag{C.52}$$

$$\tilde{\beta}^*_{n,t}(\overline{\Delta x}_{n,t} - \Delta\tilde{x}^*_{n,t}) = 0, \tag{C.53}$$

$$\tilde{\theta}^*_{n,t}\tilde{p}^*_{n,t} = 0, \tag{C.54}$$

$$\tilde{\alpha}^*_{n,t} \geq 0, \quad \tilde{\beta}^*_{n,t} \geq 0, \quad \tilde{\theta}^*_{n,t} \geq 0, \tag{C.55}$$

respectively.

To prove $\tilde{V}_{n,t}(S_t)$ is concave in x_t, note that by induction, $\tilde{V}_{n,t^+}\big(x_{t^+} + \Delta\tilde{x}_t, \mathbf{w}_t\big)$ is concave and increasing in x_{t^+}. Defining the affine function $B(\Delta\tilde{x}_t) = x_t + \Delta\tilde{x}_t$ shows $\tilde{V}_{n,t^+}$ is also concave in $\Delta\tilde{x}_t$ (See Rockafellar (2015), Theorem 5.1). Now considering two different x_t and using the same approach used in the proof of concavity of $V_{1^+}(x_1)$ in Eq. (B.10), we can show $\tilde{V}_{n,t}(S_t)$ is concave in x_t.

C.2 Proof of Results in Section 2.6.3

Proof of Theorem 9, Part (i) We prove by construction. Assume $\hat{\sigma}^*_{n,t}$ and $\hat{\gamma}^*_{n,t}$ are the corresponding strategies for the dominant SPNE of the dynamic game induced by the mechanism proposed in Section (2.6.2) . By Lemma 3, these strategies are functions of S_t and S_{t^+} respectively. Define $\hat{e}^*_{n,t}(S_{t^+})$ to be GenCo n's production at time t^+ due to strategy $\tilde{\gamma}^*_{n,t}(S_{t^+})$ and $\Delta\tilde{x}^*_{n,t}(S_t)$ to be GenCo

n's expansion at time t due to strategy $\tilde{\sigma}^*_{n,t}(S_t)$. Consider the following strategy for the planner.

$$\sigma^*_{n,t}(S_t) = \Delta\tilde{x}^*_{n,t}(S_t) \tag{C.56}$$

$$\gamma^*_{n,t}(S_{t^+}) = \hat{e}^*_{n,t}(S_{t^+}) \tag{C.57}$$

Note that the capacity expansion is equivalent to an auction with inelastic demand; therefore, using the results of Theorem (3), Lemma (1), the KKT conditions of $\tilde{Q}_{3,n}(t)$ and $\tilde{Q}_{3,n}(t^+)$, Eq. (C.50)-(C.55), Eq. (C.40)-(C.45) respectively, and the generation and expansion strategies of Eq. (C.56)-(C.57), the following parameters will satisfy the KKT conditions of $Q_3(t)$, Eq. (C.28)-(C.32), and the KKT conditions of $Q_3(t^+)$, Eq. (C.19)-(C.22). For all $t \in T$ and $n \in N$ put

$$\nu^*_{n,t} = \tilde{\nu}^*_{n,t} \tag{C.58}$$

$$\mu^*_{n,t} = \tilde{\mu}^*_{n,t} \tag{C.59}$$

$$\alpha^*_{n,t} = \tilde{\alpha}^*_{n,t} \tag{C.60}$$

$$\beta^*_{n,t} = \tilde{\beta}^*_{n,t} \tag{C.61}$$

As a result, the energy production and expansion corresponding to the non-trivial SPNE are solutions of Problem (Socially Optimal - Longterm Uncertainty).

Proof of Theorem 9, Part (ii) and (iii) To prove price efficiency and budget balanced in spot (or capacity) market of time t^+ (time t), we model this market as a single elastic (or inelastic) auction presented in Section 2.3.2 and use the properties of these auctions presented in Section 2.3.5. To do this modeling, every GenCo's cost of generation (cost of expansion) is the sum of both the immediate utility from the spot (or capacity) market and the value of the future outcomes, $C^e_{n,t}(\hat{e}_{n,t}, w_t) - \beta\tilde{V}_{n,t+1}(S_t)$, (or $C^x_{n,t}(\Delta\tilde{x}_{n,t}, w_{t-1}) - \beta E_{\mathbf{w}_t}\left[\tilde{V}_{t^+,n}\left(S_t, \Delta\tilde{x}_t, \tilde{p}_t, \mathbf{w}_t\right)\middle| w_{1:t-1}\right]$)). This cost function is strictly convex and increasing and therefore the results of market clearance, Lemma (1), price efficiency, Theorem (1), and budget balance, Theorem (2), can be generalized for the spot (or capacity) market at time t^+ (t). This means Equations (2.78), (2.79) and (2.80) hold.

Note that for the capacity market, the price at equilibrium is the dual variable of the long-term reliability constraint κ^*_t (Eq. C.27). From Equation (C.50), κ^*_t represents the difference between the marginal cost of expansion and the marginal value of the future profits if $0 < \Delta\tilde{x}_{n,t} < \overline{\Delta x}_{n,t}$ i.e. $\hat{\alpha}^*_{n,t} = 0$ and $\hat{\beta}^*_{n,t} = 0$.

Proof of Theorem 9, Part (iv) To prove individual rationality, we should prove that, at each time instant t, by continuing participation in the mechanism, the GenCos' future utility is greater than or equal to zero (which is the utility they get by stopping to participate in the mechanism). We prove this by induction. We show that at time $T+1$ the GenCos' values are non-negative and then we show that if at any time t (respectively t^+) it is weakly better for the GenCos to continue the game, then so is at $(t-1)^+$ (respectively t).

First, note that at time $T+1$, $V_{n,T+1}, n \in N$ is non-negative by definition. Now assume at time $t+1$ it is individually rational for the GenCos to participate in the mechanism. Consider the generation epoch t^+. Since at the dominant SPNE the firms act myopically in time t^+, by Theorem (4) their immediate revenue from the spot market is non-negative; therefore, the assumption that individual rationality holds at time $t+1$ results that it also holds at time t^+.

Next, assume the mechanism is individually rational starting from any S_{t^+} at time t^+. Consider time t. For the non-trivial NE, GenCo n's utility from Theorem 9 Part (i) is $-C^x_{n,t}(\tilde{\Delta x}_{n,t}) + E\{\tilde{V}_{n,t}(x_t + \Delta\tilde{x}^*_t, w_{1:t})|S_t\} + \hat{p}^*_t \Delta\tilde{x}^*_{n,t}$ where $\Delta\tilde{x}^*_t$ is the solution of $Q_3(t)$; if $0 = \Delta\tilde{x}^*_{n,t}$, then this utility is $E\{\tilde{V}_{n,t}(x_t + \Delta\tilde{x}^*_t, w_{1:t})|S_t\}$ which is non-negative by assumption. If $0 < \Delta\tilde{x}^*_{n,t}$ then

$$\hat{p}^*_t \geq \frac{\partial\Big(- C^x_{n,t}(\tilde{\Delta x}_{n,t}) + E\{\tilde{V}_{n,t}(x + \Delta\tilde{x}^*_t, w_{1:t}|S_t)\}\Big)}{\partial \Delta\tilde{x}^*_{n,t}}; \tag{C.62}$$

since $-C^x_{n,t}(\tilde{\Delta x}_{n,t}) + E\{\tilde{V}_{n,t}(x + \Delta\tilde{x}^*_t, w_{1:t})|S_t\}$ is strictly concave and the price $\hat{p}^*_t$ is the marginal of this utility, the participation utility at equilibrium $-C^x_{n,t}(\tilde{\Delta x}_{n,t}) + E\{\tilde{V}_{n,t}(x_t + \Delta\tilde{x}^*_t, w_{1:t})|S_t\} + \hat{p}^*_t \Delta\tilde{x}^*_{n,t}$ is non-negative. Therefore, it is individually rational to continue participation in the game at time t.

Remark C.2.1. *Note that since Problem (Socially Optimal - Longterm Uncertainty) has a unique solution, its production amounts and marginal cost of productions uniquely determine the non-trivial SPNE. Moreover, in the capacity market of time $t \in T$ for every history $\tilde{h}_t \in \tilde{H}_t$, $t \in T$, GenCos can choose between the trivial and the non-trivial NE strategies. Therefore, there is a set of SPNEs for the game induced by the mechanism. However, the unique non-trivial SPNE in which GenCos choose the non-trivial stage NE at all histories $h_t \in H_t, \forall t \in T$ Pareto dominates other SPNEs. Intuitively, this is because if GenCos expand their capacity, then they are compensated at the marginal cost of expansion in the capacity market. Since the expansion cost is convex, GenCos collect profit from the capacity markets.*

Proof of Theorems 10 Consider the problem of socially optimal expansion and generation with infinite horizon. At the optimal Markov strategy, the value function $V(S), V^{+,*}(S^+)$ should satisfy the Bellman equations:

$$V(S) = \max_{\Delta x_n} \big\{ \sum_{n \in N} -C^x_n(\Delta x_n, w) + E_{\mathbf{w}'|w}\{V^{+,*}(S^+)|w\}\big\}, \tag{C.63}$$

where

$$S^+ = (x(S) + \Delta x, w'), \tag{C.64}$$

and

$$V^{+,*}(S^+) = \max_{e_n} \big\{ \sum_{n \in N} -C_n(e_n, w) + U(\sum_{n \in N} e_n, w) + \beta V(S)\big\}, \tag{C.65}$$

where

$$S = S^+. \tag{C.66}$$

These equations form a contraction mapping and therefore, they have a unique solution V and $V^{+,*}$ that satisfy (C.63) and (C.65).

GenCos' best responses satisfy the following Bellman equations (note that for the infinite horizon we have assumed the random noise $w_{1:T}$ is Markov).

$$\hat{V}_n^*(S) = \max_{\Delta\tilde{x}_n, \tilde{p}_n} \left\{ -C_n^x(\Delta\tilde{x}_n, w) + r_n^{inelas.}(\Delta\tilde{x}_n, \tilde{p}_n, \Delta\tilde{x}_{-n}, \tilde{p}_{-n}) + E_{\mathbf{w}'|w}\{\hat{V}_n^{+*}(S^+)\} \right\} \tag{C.67}$$

$$S^+ = (x(S) + \Delta\tilde{x}, w') \tag{C.68}$$

and

$$\hat{V}_n^{+,*}(S^+) = \max_{\hat{e}_n, \hat{p}_n} \left\{ -C_n(\hat{e}_n, w) + r_n(\hat{e}_n, \hat{p}_n, \hat{e}_{-n}, \hat{p}_{-n}) + \beta V(S) \right\}, \tag{C.69}$$

$$S = S^+ \tag{C.70}$$

Construct the strategies and the value functions of Problem (Socially Optimal - Longterm Uncertainty) form strategies and value functions corresponding to the dominant MPE of the game induced by the mechanism. Define the strategies of Problem (Socially Optimal - Longterm Uncertainty) by Eq. (C.56)-(C.57) and value functions by

$$\begin{aligned} V(S) = E_{\mathbf{w}_{2:\infty}|w}\{ &\sum_{t\in\{2,3,\ldots\}, n\in N} \beta^t \left\{ -C_n^x(\Delta\tilde{x}_{n,t}^*, \mathbf{w}_{t-1}) - C_n^e(\hat{e}_{n,t}^*, w_t) \right\} \\ &+ \sum_{t=2,3,\ldots,\infty} U'(\sum_{m\in N} \hat{e}_{m,t}^*, w_t)\hat{e}_{n,t}^* \} \end{aligned} \tag{C.71}$$

$$\begin{aligned} V_n^{+,*}(S^+) = &\sum_{n\in N} -C_n^e(\hat{e}_{n,1}^*, w) + U'(\sum_{m\in N} \hat{e}_{m,1}^*, w)\hat{e}_{n,t}^* \\ &+ E_{\mathbf{w}_{2:\infty}|w}\{ \sum_{n\in N, t\in\{2,3,\ldots\}} \beta^t \Big(-C_n^x(\Delta\tilde{x}_{n,t}^*, \mathbf{w}_t) \\ &- C_n^e(\hat{e}_{n,t}^*, \mathbf{w}_{t+1}) + U'(\sum_{m\in N} \hat{e}_{m,t}^*, \mathbf{w}_{t+1})\hat{e}_{n,t}^* \Big) \} \end{aligned} \tag{C.72}$$

These strategies along with value functions $V_n^*(S)$ and $V_n^{+,*}(s^+)$ satisfy the Bellman equations for Problem (Socially Optimal - Longterm Uncertainty), Eqs. (C.63) and (C.65).

APPENDIX D

Extension of the Auctions with Individual Constraints in Section (2.4) to Long-term Interaction of GenCos with Symmetric Uncertainty

D.1 Model

We extend the model for investment without reliability constraints of Section (2.4) to the case where GenCos interact over a finite time horizon and have symmetric uncertainty that is common knowledge among them. The expected social welfare maximizing investment and production are given by the solution of the following problem

$$\max_{\sigma_{n,t}\in\Sigma_{n,t},\gamma_{n,t}\in\Gamma_{n,t}:n\in N,t\in T} E_{\mathbf{w}_{1:T}}\Bigg[\sum_{t\in T}\beta^t U_t(\sum_{n\in N} e_{n,t},\mathbf{w}_t)- \qquad \text{(Socially Optimal - Longterm Uncertainty II)}$$

$$\sum_{t\in T}\sum_{n\in N}\beta^t\Big\{C^e_{n,t}(e_{n,t},\mathbf{w}_t)+C^x_{n,t}(\Delta x_{n,t},\mathbf{w}_{t-1})\Big\}+\beta^{T+1}\sum_{n\in N}\eta(\mathbf{w}_T)x_{n,T}\Bigg] \tag{D.1}$$

$$s.t.\quad 0\le e_{n,t}\le x_{n,0}+\sum_{\tau=1}^{t}\Delta x_{n,\tau}\quad \forall t\in T \tag{D.2}$$

$$0\le \Delta x_{n,t}\le \overline{\Delta x}_{n,t}(\mathbf{w}_{t-1})\quad \forall t\in T \tag{D.3}$$

In Lemma (4), we prove that at any time t, without loss of optimality we can restrict attention to expansion strategies $\sigma^*_{n,t}$ that are functions of $S_t=\{x_{t-1},w_{1:t-1}\}$ and to generation strategies, $\gamma^*_{n,t}$ that are only functions of $S_{t^+}=\{x_t,w_{1:t}\}$. In other words, S_t and S_{t^+} are information states for Problem (Socially Optimal - Longterm Uncertainty). We also prove that the same is true for the GenCo's SPNE strategies in the game induced by the proposed mechanism.

D.2 Design/Mechanism

Using dynamic programming, Problem (Socially Optimal - Longterm Uncertainty II) can be decomposed into a set of $2\times T$ sequential optimization problems, two for each time step (See Section (D.4) for details). Considering this decomposition, we propose the following dynamic market mechanism for implementing the outcome. During each year $t\in T$, first the GenCos individually decide on expanding their capacities at time t. Next, at time t^+, they participate in the efficient auction (spot market) for elastic demand $U_t(d_t, w_t)$ proposed in Section (2.3.2).

D.3 Properties of the Mechanism

At each time $t \in T$, GenCo n's, $n \in N$, strategy consists of an expansion strategy $\hat{\sigma}_{n,t}$ and a production strategy $ga\hat{m}ma_{n,t}$ that describes how he participates in the spot market. The history of the game at t, the time of expansion is

$$\hat{h}_t = \{x_0, \Delta\hat{x}_{1:t-1}, \hat{e}_{1:t-1}, \hat{p}_{1:t-1}, w_{1:t-1}\} \in \hat{H}_t \tag{D.4}$$

$$\hat{H}_t := \{x_0, \Delta\hat{X}_{1:t-1}, \hat{E}_{1:t-1}, \hat{P}_{1:t-1}, W_{1:t-1}\}, \tag{D.5}$$

and the GenCos expansion strategies are

$$\hat{\sigma}_{n,t} : \hat{H}_t \to [0, \overline{\Delta x}_{n,t}], n \in N; \tag{D.6}$$

the history of the game at t^+, the time of participation in the spot market, is

$$\hat{h}_{t^+} = \{x_0, \Delta\hat{x}_{1:t}, \hat{e}_{1:t-1}, \hat{p}_{1:t-1}, w_{1:t}\} \in \hat{H}_{t^+} \tag{D.7}$$

$$\hat{H}_{t^+} = \{x_0, \Delta\hat{X}_{1:t}, \hat{E}_{1:t-1}, \hat{P}_{1:t-1}, W_{1:t}\}, \tag{D.8}$$

and the GenCos production strategies are,

$$\hat{\gamma}_{n,t^+} : \hat{H}_{t^+} \to [0, x_{n,t}] \times \mathbb{R}_+ \tag{D.9}$$

We denote the set of all expansion and production strategies by $\hat{\Sigma}_{n,t}$ and $\hat{\Gamma}_{n,t}$, respectively.

We first show that in the finite horizon economy, the producers SPNE expansion and production strategies are Markov i.e. they are only a function of the existing capacity and the realized uncertainties (Lemma (4)). We prove that this feature of equilibrium strategies implies that GenCos are myopic in the spot market. We use this to prove the sub-game perfect Nash equilibrium of the game induced by the mechanism satisfies market clearing, price efficiency and budget balance (Theorem (13)). We also prove social optimality, individual rationality, and uniqueness of the equilibrium in Theorems (14), (15) and (16) correspondingly. Theorem (17) shows that the GenCos cover their cost of expansion i.e. there is no missing money problem.

Lemma 4. *(MARKOV STRATEGIES) For Problem (Socially Optimal - Longterm Uncertainty II), there exists an optimal expansion strategy at time t and optimal production strategy at time t^+ which are functions of $S_t := \{x_{t-1}, w_{1:t-1}\}$ and $S_{t^+} = \{x_t, w_{1:t}\}$, $t \in T$ respectively.*

Also, for the game induced by the mechanism of Section (D.2), any SPNE is a function of S_t at time t and S_{t^+} at time t^+.

Lemma 4 implies that GenCos behave myopically in the spot markets i.e. maximize their immediate revenue from the spot market. This is because the actions in the spot market i.e. production and price bids, do not change the information state of the system for future Markov strategies i.e. existing capacity and realized uncertainties[1]. Therefore, the price and production bids

[1]We will later show that the equilibrium is unique and there is not an equilibrium selection issue.

will be similar to those of a single elastic spot market studied in Section (2.3.5). This discussion provides the intuition for the feasibility, budget balanced and price efficiency of the equilibrium as presented in the Theorem below.

Theorem 13 (MARKET CLEARANCE, PRICE EFFICIENCY, BUDGET BALANCED)**.** *For every time $t \in T$ and for every realization of the history $\hat{h}_{t^+} \in \hat{H}_{t^+}$, the game induced by the mechanism proposed in Section (D.2) has the following properties at its SPNE: The outcome of spot market is feasible, the electricity production price is efficient, and the payments are budget balanced, i.e. $\forall t \in T$*

$$\hat{p}^*_{n,t} = \hat{p}^*_t = C'(e^*_{n,t}) \text{ if } 0 < e^*_{n,t} < x_{n,t}, \tag{D.10}$$

$$\hat{p}^*_t(\hat{h}_{t^+}) = U'_t(\sum_{n \in N} e^*_{n,t}), \tag{D.11}$$

$$\sum_{n \in N} r^*_{n,t}(\hat{h}_{t^+}) = \hat{p}^*_t(\hat{h}_{t^+}) \times \sum_{n \in N} e^*_{n,t}(\hat{h}_{t^+}). \tag{D.12}$$

Theorem 14 (SOCIAL OPTIMALITY)**.** *The expansion and production outcomes corresponding to every SPNE are socially optimal, that is they are equal to the solution of Problem (Socially Optimal - Longterm Uncertainty II).*

Theorem 15 (INDIVIDUAL RATIONALITY)**.** *The SPNE of the game induced by the mechanism is individually rational. That is at each stage, the future expected allocations corresponding to the SPNE of the game induced by the mechanism are weakly preferred by all producers to the future allocation $(0,0)$ (the allocation a producer receives when it does not participate anymore in the mechanism)*[2].

Theorem 16 (UNIQUENESS OF SPNE)**.** *The game induced by the proposed mechanism of Section (D.2), has a unique SPNE.*

Theorem 17 (SATURATION)**.** *Let $(\hat{\sigma}^*_{n,t}, \hat{\gamma}^*_{n,t}, n \in N, t \in T)$ denote the SPNE of the game induced by the mechanism. If $\hat{\sigma}^*_{n,t}(\hat{h}_{n,t}) > 0$ then there exists a future time instant $\tau \geq t$ and a continuation $\hat{h}_{t+1:\tau}$ of $\hat{h}_t$ such that $p(\hat{h}_t, \hat{h}_{t+1:\tau} | \hat{h}_t) > 0$ and $\hat{\gamma}^*_{n,t}(\hat{h}_{\tau^+})$ is equal to GenCo n's capacity $\hat{x}^*_{n,\tau}$.*

When a GenCo is saturated, the market price is greater or equal to its marginal cost of production. This extra income covers the GenCo's cost of expansion (no missing money). Therefore, existence of saturation under at least one scenario with positive probability is necessary for accepting the cost for expanding the capacity and yet having a positive expected future revenue.

D.4 Preliminaries for Proofs

The solution to Problem (Socially Optimal - Longterm Uncertainty II) and the determination of SPNE can be obtained by backward induction. We present the corresponding stage optimization problems at times t and t^+, denoted by $Q_2(t)$ & $Q_2(t^+)$. We also present GenCo n's best response

[2] Note that there may exist a history $\hat{h}_t$ and GenCo $n \in N$ such that the accumulated revenue from the past is negative. Even more, there may be a history and a GenCo such that the sum of his accumulated past revenues and his future expected revenue is negative.

problem at times t and t^+, $\hat{Q}_{2,n}(t)$ and $\hat{Q}_{2,n}(t^+)$. To show these are strictly concave optimization problems, we prove Lemma (4). We present the Lagrangians and KKT conditions for these stage problems.

Problem (Socially Optimal - Longterm Uncertainty II)

For social optimality, Problem (Socially Optimal - Longterm Uncertainty II), we define the stage problems $Q_2(t)$ & $Q_2(t^+)$ respectively.

Let

$$V_{T+1}(h_{T+1}) := \eta(w_T) \sum_{n \in N} x_{n,T+1}. \tag{D.13}$$

$Q_2(t^+)$

At time t^+ optimal generations are solution to the following optimization problem.

$$\begin{aligned} V_{t^+}(h_{t^+}) = & \max_{e_{n,t}, n \in N} \sum_{n \in N} U_t(\sum_{n \in N} e_{n,t}, w_t) - \sum_{n \in N} C^e_{n,t}(e_{n,t}, w_t) + \beta V_{t+1}(h_{t^+}, e_t) && \text{(D.14)} \\ & s.t. \quad 0 \leq e_{n,t} \leq x_{n,t^+} \forall n \in N && \text{(D.15)} \end{aligned}$$

Note that $(h_{t^+}, e_t) = h_{t+1}$.

$Q_2(t)$

The expansions at time t are solution to the following optimization problem.

$$\begin{aligned} V_t(h_t) = & \max_{\Delta x_{n,t}, n \in N} -\sum_n C^x_{n,t}(\Delta x_{n,t}, w_{t-1}) + \beta E_{\mathbf{w}_t}\left[V_{t^+}(h_t, \Delta x_t, \mathbf{w}_t) | w_{1:t-1}\right] && \text{(D.16)} \\ & s.t. \quad 0 \leq \Delta x_{n,t} \leq \overline{\Delta x}_{n,t} && \text{(D.17)} \end{aligned}$$

Note that $(h_t, \Delta x_t, w_t) = h_{t^+}$.

GenCos' Best Response

For GenCo n, $n \in N$ we define best response problems $\hat{Q}_{2,n}(t)$ and $\hat{Q}_{2,n}(t^+)$ at time t and t^+ respectively. The history $\hat{h}_t$ at time t (resp. the history $\hat{h}_{t^+}$ at time t^+) is an information state for problem $\hat{Q}_2(n,t)$ (respectively problem $\hat{Q}_{2,n}(t^+)$).

First consider

$$\hat{V}_{T+1,n}(\hat{h}_{T+1}) := \eta(w_T) x_{n,T+1}. \tag{D.18}$$

$\hat{Q}_{2,n}(t^+)$

Consider GenCo $n \in N$. Given $\hat{m}_{-n,t} = (\hat{e}_{-n,t}, \hat{p}_{-n,t})$ to be the other GenCos' strategies at time t^+, GenCo n's best response at the spot market of time t^+ with history $\hat{h}_{t^+}$ is a solution of

$$\begin{aligned} & \max_{\hat{e}_{n,t}, \hat{p}_{n,t}} - C^e_{n,t}(\hat{e}_{n,t}) + \hat{p}_{n+1,t}\hat{e}_{n,t} - \hat{p}_{n,t}^{-0.5} \zeta_{n,t}^{elas.2} + \beta \hat{V}_{n,t+1}(\hat{h}_{t^+}, \hat{p}_t, \hat{e}_t) && \text{(D.19)} \\ & s.t. \quad 0 \leq \hat{e}_{n,t} \leq x_{n,t} && \text{(D.20)} \\ & \qquad\quad 0 \leq \hat{p}_{n,t}. && \text{(D.21)} \end{aligned}$$

$\hat{Q}_{2,n}(t)$

GenCo n's best response for expansion decision at time t, Given the other GenCo's expansion strategy to be $m_{-n,t} = \Delta\hat{x}_{-n,t}$, is

$$\hat{V}_{n,t}(\hat{h}_t) = \max_{\Delta\hat{x}_{n,t}} -C^x_{n,t}(\Delta\hat{x}_{n,t}, w_{t-1}) + \beta E_{\mathbf{w}_t}\left[\hat{V}_{n,t^+}\left(\hat{h}_t, \Delta\hat{x}_t, \mathbf{w}_t\right)\Big| w_{1:t-1}\right] \tag{D.22}$$

$$0 \leq \Delta\hat{x}_{n,t} \leq \overline{\Delta x}_{n,t}, \tag{D.23}$$

To prove strict concavity of the above stage optimization problems, we prove Lemma (4).

Proof of Lemma 4 We show that $S_t := (x_{t-1}, w_{1:t-1})$ (respectively $S_{t^+} := (x_t, w_{1:t})$) is an information state for Problem (Socially Optimal - Longterm Uncertainty II) and then for the SPNE strategies of the game induced by the mechanism. We proceed by induction.

Problem (Socially Optimal - Longterm Uncertainty II)

We proceed in three steps. We first show that the value function at time $T+1$, V_{T+1} is a function of S_{T+1}. Then we prove that if for $t+1 \in T$, V_{t+1} is function of S_{t+1} then V_{t^+} is also a function of S_{t^+}. Finally, we show that if for $t \in T$, V_{t^+} is a function of S_{t^+}, then V_t is a function of S_t.

First note that by (Eq. D.13), at time $T+1$, V_{T+1} is a function of S_{T+1}.

Next, assume V_{t+1} is a function of S_{t+1} at equilibrium, and therefore independent of $e_{n,t}$. Then, Problem $Q_2(t^+)$ is a function of S_{t^+} because in the optimization in Eqs. (D.14)-(D.15), Eq. (D.15) is a function of x_{t^+}, and V_{t+1} is a function of S_{t^+}.

Finally, assume V_{t^+} is only a function of S_{t^+}. Then Problem $Q_2(t)$ is a function of S_t.

GenCos' Best Response Problem

We proceed again in three steps. We first show that the value function at time $T+1$, $\hat{V}_{n,T+1}$ is a function of S_{T+1}. Then we prove that if for $t+1 \in T$, $\hat{V}_{n,t+1}$ is function of S_{t+1} for all $n \in N$ then $\hat{V}_{n,t^+}$ is also a function of S_{t^+} for all $n \in N$. Finally, we show that if for $t \in T$, $\hat{V}_{n,t^+}$ is a function of S_{t^+} for all $n \in N$, then $\hat{V}_{n,t}$ is a function of S_t for all $n \in N$.

First note that by definition, at time $T+1$, $\hat{V}_{n,T+1}$ is a function of S_{T+1} (Eq. D.18).

Next, assume $\hat{V}_{n,t+1}$ is a function of S_{t+1} for all $n \in N$. It is straightforward to see that as a result, Problem $\hat{Q}_{2,n}(t^+)$ is a function of S_{t^+}.

Finally, assume $\hat{V}_{n,t^+}$ is a function of S_{t^+} for all $n \in N$. Since S_{n,t^+} is determined by S_t and the expansion at time t, $\Delta\hat{x}_{n,t}$, Problem $\hat{Q}_{2,n}(t)$ and its constraints can be written as a function of S_t. As a result the strategies at time t are function of S_t[3].

Using the above lemma, we present the Lagrangian and KKT conditions.

Problem (Socially Optimal - Longterm Uncertainty II)

For social optimality, Problem (Socially Optimal - Longterm Uncertainty II), we define the stage problems $Q_2(t)$ & $Q_2(t^+)$ respectively.

Let

$$V_{T+1}(S_{T+1}) := \eta(w_T) \sum_{n \in N} x_{n,T+1}. \tag{D.24}$$

$V_{T+1}(S_{T+1})$ is concave in x_t.

$Q_2(t^+)$

[3] There is not an equilibrium selection problem for the uniqueness of the equilibrium.

At time t^+ optimal generations are solution to the following optimization problem.

$$V_{t^+}(S_{t^+}) = \max_{e_{n,t}, n \in N} \sum_{n \in N} U_t(\sum_{n \in N} e_{n,t}, w_t) - \sum_{n \in N} C^e_{n,t}(e_{n,t}, w_t) + \beta V_{t+1}(S_{t^+}) \tag{D.25}$$

$$s.t. \quad 0 \le e_{n,t} \le x_{n,t^+} \forall n \in N \tag{D.26}$$

Note that $S_{t+1} = S_{t^+}$. Since $V_{t+1}(S_{t^+})$ is constant with respect to $e_{n,t}$, the above optimization is strictly concave with convex domain. The corresponding Lagrangian and KKT conditions are

$$\begin{aligned} L_{t^+}(S_{t^+}) = U_t(\sum_{n \in N} e_{n,t}, w_t) - \sum_{n \in N} C^e_{n,t}(e_{n,t}, w_t) + \beta V_{t+1}(S_{t^+}) \\ + \nu_{n,t} e_{n,t} + \mu_{n,t}(x_{n,t^+} - e_{n,t}). \end{aligned} \tag{D.27}$$

and

$$\begin{aligned} \frac{\partial U_t(\sum_{n \in N} e_{n,t}, w_t)}{\partial e_{n,t}}\Big|_{e^*_t} - \frac{\partial C^e_{n,t}(e_{n,t}, w_t)}{\partial e_{n,t}}\Big|_{e^*_t} \\ + \beta \frac{\partial V_{t+1}(S_{t^+})}{\partial e_{n,t}}\Big|_{e^*_t} + \nu^*_{n,t} - \mu^*_{n,t} = 0 \end{aligned} \tag{D.28}$$

$$\nu^*_{n,t} e^*_{n,t} = 0 \tag{D.29}$$

$$\mu^*_{n,t}(x_{n,t^+} - e^*_{n,t}) = 0 \tag{D.30}$$

$$\nu^*_{n,t} \ge 0, \quad \mu^*_{n,t} \ge 0, \tag{D.31}$$

respectively.

To prove $V_{t^+}(S_{t^+})$ is concave with respect to x_{t^+} for every $w_{1:t} \in W_{1:t}$, it is straightforward to extend the same proof approach used for concavity of $V_{t^+}(S_{t^+})$ in Eq. (C.15) to $V_{t^+}(S_{t^+})$.

$\underline{Q_2(t)}$

The expansions at time t are solution to the following optimization problem.

$$V_t(S_t) = \max_{\Delta x_{n,t}, n \in N} -\sum_n C^x_{n,t}(\Delta x_{n,t}, w_{t-1}) + \beta E_{\mathbf{w}_t}\Big[V_{t^+}(x_t + \Delta x_t, \mathbf{w}_{1:t}) | w_{1:t-1}\Big] \tag{D.32}$$

$$s.t. \quad 0 \le \Delta x_{n,t} \le \overline{\Delta x}_{n,t} \tag{D.33}$$

Note that $(x_t + \Delta x_t, w_{1:t}) = S_{t^+}$. Assuming that $V_{t^+}(x_t + \Delta x_t, \mathbf{w}_{1:t})$ is concave with respect to x_t (by induction), we prove $Q_3(t)$ is strictly concave optimization problem and furthermore, $V_t(S_t)$ is concave in x_t. First, to prove concavity with respect to Δx_t. Define the function $B(\Delta x_t) = x_t + \Delta x_t$; $B(\Delta x_t)$ is convex and non-decreasing and therefore, $V_{t^+}(B(\Delta x_t))$ is concave (See Rockafellar (2015), Theorem 5.1). Since $E_{\mathbf{w}_t}\Big[V_{t^+}(x_{t^+} + \Delta x_t, \mathbf{w}_{1:t}) | w_{1:t-1}\Big]$ is a weighted sum of concave functions of Δx_t, it is concave in Δx_t. Considering this, it is straightforward to see that Eqs. (D.32)-(D.33) is a concave optimization problem with convex domain. The corresponding Lagrangian and KKT conditions are

$$L_t(S_t) = -\sum_n C^x_{n,t}(\Delta x_{n,t}, w_{t-1}) + \beta E_{\mathbf{w}_t}\left[V_{t^+}(x_t + \Delta x_t, \mathbf{w}_{1:t})|w_{1:t-1}\right] +$$
$$\sum_{n\in N}\left(\alpha_{n,t}\Delta x_{n,t} + \beta_{n,t}(\overline{\Delta x}_{n,t} - \Delta x_{n,t})\right) \quad (D.34)$$

and

$$-\frac{\partial C^x_{n,t}(\Delta x^*_{n,t}, w_{t-1})}{\Delta x^*_{n,t}} + \beta E_{\mathbf{w}_t}\left[\frac{V^*_{t^+}(x_t + \Delta x^*_t, \mathbf{w}_{1:t})}{\Delta x^*_{n,t}}|w_{1:t-1}\right] + \alpha^*_{n,t} - \beta^*_{n,t} = 0 \quad (D.35)$$
$$\alpha^*_{n,t}\Delta x^*_{n,t} = 0 \quad (D.36)$$
$$\beta^*_{n,t}(\overline{\Delta x}^*_{n,t} - \Delta x^*_{n,t}) = 0 \quad (D.37)$$
$$\alpha^*_{n,t} \geq 0, \quad \beta^*_{n,t} \geq 0 \quad (D.38)$$

To prove $V_t(S_t)$ is concave in x_t for a given $w_{1:t-1} \in W_{1:t-1}$, assume by induction that V_{t^+} is concave and follow the same approach used for proof of concavity of $V_{t^+}(S_{t^+})$ in Eq. (C.23).

GenCos' Best Response

For GenCo n, $n \in N$ we define best response problems $\hat{Q}_{2,n}(t)$ and $\hat{Q}_{2,n}(t^+)$ at time t and t^+ respectively. With these information states, we proceed to specify problems $\hat{Q}_{2,n}(t)$ and $\hat{Q}_{2,n}(t^+)$, $n \in N, t \in T$. For each problem we also define its Lagrangian and the corresponding KKT conditions.

First consider

$$\hat{V}_{T+1,n}(S_{T+1}) := \eta(w_T)x_{n,T+1}. \quad (D.39)$$

Note that $\hat{V}_{T+1}(S_{T+1})$ is concave in x_{T+1}.

$\hat{Q}_{2,n}(t^+)$

Consider GenCo $n \in N$. Given $\hat{m}_{-n,t} = (\hat{e}_{-n,t}, \hat{p}_{-n,t})$ to be the other GenCos' strategies at time t^+, GenCo n's best response at the spot market of time t^+ with history S_{t^+} is a solution of

$$\hat{V}_{n,t^+}(S_{t^+}) = \max_{\hat{e}_{n,t},\hat{p}_{n,t}} -C^e_{n,t}(\hat{e}_{n,t}) + \hat{p}_{n+1,t}\hat{e}_{n,t} - \hat{p}^{-0.5}_{n,t}\zeta^{elas.^2}_{n,t} + \beta\hat{V}_{n,t+1}(S_{t+1}) \quad (D.40)$$
$$s.t. \quad 0 \leq \hat{e}_{n,t} \leq x_{n,t} \quad (D.41)$$
$$0 \leq \hat{p}_{n,t}. \quad (D.42)$$

Since S_t is constant with respect to $\hat{e}_{n,t}, \hat{p}_{n,t}$, it is also concave with respect to $(\hat{e}_{n,t}, \hat{p}_{n,t})$. Therefore, $\hat{V}_{n,t}$ is concave optimization with convex domain. The corresponding Lagrangian and KKT conditions are

$$\hat{L}_{n,t^+}(S_{t^+}) = -C^e_{n,t}(\hat{e}_{n,t}) + \hat{p}_{n+1,t}\hat{e}_{n,t} - \hat{p}^{-0.5}_{n,t}\zeta^{elas.^2}_{n,t} + \beta\hat{V}_{t+1}(S_t) + \hat{\nu}_{n,t}\hat{e}_{n,t} \quad (D.43)$$
$$+\hat{\mu}_{n,t}(x_{n,t} - \hat{e}_{n,t}) + \hat{\theta}_{n,t}\hat{p}_{n,t}, \quad (D.44)$$

and

$$-\frac{\partial C^e_{n,t}}{\partial \hat{e}_{n,t}}|_{e^*_{n,t}} + \hat{p}^*_{n+1,t} - \frac{\partial \hat{p^*}_{n,t}^{-0.5} \zeta^{elas.2}_{n,t}}{\partial \hat{e}_{n,t}}|_{(\hat{m}_{-n,t},\hat{e}^*_{n,t},\hat{p}^*_{n,t})}$$

$$+\beta \frac{\partial \hat{V}_{n,t+1}(S_{t^+})}{\partial \hat{e}_{n,t}}|_{(\hat{m}_{-n,t},\hat{e}^*_{n,t},\hat{p}^*_{n,t})} + \hat{\nu}^*_{n,t} - \hat{\mu}^*_{n,t} = 0 \tag{D.45}$$

$$-\frac{\partial \hat{p}_{n,t}^{-0.5} \zeta^{elas.2}_{n,t}}{\partial \hat{p}_{n,t}}|_{(\hat{m}_{-n,t},\hat{e}^*_{n,t},\hat{p}^*_{n,t})} + \beta \frac{\partial \hat{V}_{n,t+1}(S_{t^+})}{\partial \hat{p}_{n,t}}|_{(\hat{m}_{-n,t},\hat{e}^*_{n,t},\hat{p}^*_{n,t})} + \hat{\theta}^*_{n,t} = 0, \tag{D.46}$$

$$\hat{\theta}^*_{n,t}\hat{p}^*_{n,t} = 0, \tag{D.47}$$

$$\hat{\mu}^*_{n,t}(x_{n,t} - \hat{e}^*_{n,t}) = 0, \tag{D.48}$$

$$\hat{\nu}^*_{n,t}\hat{e}^*_{n,t} = 0, \tag{D.49}$$

$$\hat{\nu}^*_{n,t} \geq 0, \quad \hat{\mu}^*_{n,t} \geq 0, \quad \hat{\theta}^*_{n,t} \geq 0, \tag{D.50}$$

respectively.

To prove that for every given $w_{1:t}$, $\hat{V}_{n,t^+}(S_{t^+})$ is concave with respect to x_{n,t^+}, assume $\hat{V}_{n,t+1}$ is concave in x_{t+1} for every $w_{1:t} \in W_{1:t}$ (by induction). Then, it is straightforward to extend the same approach used for proof of concavity of $\hat{V}_{n,1^+}$ in Eq. B.30 for $\hat{V}_{n,t^+}(S_{t^+})$.

$\underline{\hat{Q}_{2,n}(t)}$

GenCo n's best response for expansion decision at time t, Given the other GenCo's expansion strategy to be $\hat{m}_{-n,t} = \Delta\hat{x}_{-n,t}$, is

$$\hat{V}_{n,t}(S_t) = \max_{\Delta\hat{x}_{n,t}} -C^x_{n,t}(\Delta\hat{x}_{n,t}, w_{t-1}) + \beta E_{\mathbf{w}_t}\left[\hat{V}_{n,t^+}\big(x_t + \Delta\hat{x}_t, \mathbf{w}_{1:t}\big)\big|w_{1:t-1}\right] \tag{D.51}$$

$$0 \leq \Delta\hat{x}_{n,t} \leq \overline{\Delta x}_{n,t}, \tag{D.52}$$

Assume (by induction) that $\hat{V}_{n,t^+}(x_t + \Delta\hat{x}_t, w_{1:t})$ is concave in $x_t + \Delta\hat{x}_t$. This means it is also concave in $\Delta\hat{x}$ (See Rockafellar (2015), Theorem 5.1). We show that the above optimization is concave optimization and furthermore, $\hat{V}_{n,t}(S_t)$ is concave in x_t for every $w_{1:t} \in W_{1:t}$. First, $E_{\mathbf{w}_{1:t}}\left[V_{n,t^+}(x_{t^+} + \Delta x_t, \mathbf{w}_t)|w_{1:t-1}\right]$ is a weighted sum of concave functions hence concave. Considering this, it is straightforward to see $V_{n,t}(S_t)$ is a strictly concave optimization with convex domain. The corresponding Lagrangian and KKT conditions are

$$\hat{L}_{n,t}(S_t) = -C^x_{n,t}(\Delta\hat{x}_{n,t}, w_{t-1}) + \beta E_{\mathbf{w}_t}\left[\hat{V}_{t^+,n}\big(x_t + \Delta\hat{x}_t, \mathbf{w}_{1:t}\big)\big|w_{1:t-1}\right] \tag{D.53}$$

$$+\hat{\alpha}_{n,t}\Delta\hat{x}_{n,t} + \hat{\beta}_{n,t}(\overline{\Delta x}_{n,t} - \Delta\hat{x}_{n,t}), \tag{D.54}$$

and

$$-\frac{\partial C^x_{n,t}(\Delta\hat{x}_{n,t},\mathbf{w}_{t-1})}{\partial \Delta\hat{x}_{n,t}}|_{(\hat{m}_{-n,t},\hat{e}^*_{n,t},\hat{p}^*_{n,t})}$$

$$+\beta E_{\mathbf{w}_t}\left[\frac{\partial \hat{V}_{n,t^+}\left(x_t+\Delta\hat{x}_t,\mathbf{w}_{1:t}\right)}{\partial \Delta\hat{x}_{n,t}}|_{\Delta\hat{x}^*_t}\middle| w_{1:t-1}\right]+\hat{\alpha}^*_{n,t}-\hat{\beta}^*_{n,t}=0, \tag{D.55}$$

$$\hat{\alpha}^*_{n,t}\Delta\hat{x}^*_{n,t}=0, \tag{D.56}$$

$$\hat{\beta}^*_{n,t}(\overline{\Delta x}_{n,t}-\Delta\hat{x}^*_{n,t})=0, \tag{D.57}$$

$$\hat{\alpha}^*_{n,t}\geq 0,\quad \hat{\beta}^*_{n,t}\geq 0. \tag{D.58}$$

respectively.

To prove $\hat{V}_{n,t}(S_t)$ is concave in x_t. First assume (by induction) that $\hat{V}_{n,t^+}$ is concave with respect to $x_t+\Delta\hat{x}_t$; therefore, it is also concave with respect to x_t (See Rockafellar (2015), Theorem 5.1). Now considering two different x_t and using the same approach used in the proof of concavity of $V_{1^+}(x_1)$ in Eq. (B.10), we can show $\hat{V}_{n,t}(S_t)$ is concave in x_t.

D.5 Proof of the Results in Section D.3

Proof of Theorem 13 Due to Lemma 4, the SPNE strategies in the spot markets are myopic for maximizing immediate revenue from the spot market. This is because the strategies in the spot market at time t^+ do not change the state S_τ for any $\tau > t^+$. As a result, each spot market can be studied as an independent static market without considering the future continuation valuation. Therefore, market clearance, budget balance and price efficiency of each spot market is independent of the other markets, and can be established in a similar way to the properties proved for Problem (Socially Optimal Dispatch - Elastic) in Lemma (1), Theorems (1) and (2).

Proof of Theorem 14 Assume $\hat{\sigma}^*_{n,t}$ and $\hat{\gamma}^*_{n,t^+}$ form a SPNE for the dynamic game induced by the mechanism proposed in Section (D.2). By Lemma (4), these strategies are functions of S_t and S_{t^+} receptively which are subsets of h_t and h_{t^+}, the history in Problem (Socially Optimal - Longterm Uncertainty II). Define $\hat{e}^*_{n,t}(S_{t^+})$ to be GenCo n's production at time t^+ due to strategy $\hat{\gamma}^*_{n,t}(S_{t^+})$ and $\Delta\hat{x}^*_{n,t}(S_t)$ to be GenCo n's expansion at time t due to strategy $\hat{\sigma}^*_{n,t}(S_t)$. Consider the following strategies in Problem (Socially Optimal - Longterm Uncertainty II).

$$\sigma^*_{n,t}(S_t)=\Delta\hat{X}^*_{n,t}(S_t), \tag{D.59}$$

$$\gamma^*_{n,t}(S_{t^+})=\hat{e}^*_{n,t}(S_{t^+}), \tag{D.60}$$

then using the results of Theorem (3), Eq. (D.55)-(D.58), Eqs. (D.59)-(D.60) and Eqs. (D.45)-(D.50), the following parameters, defined for all $t\in T$ and $n\in N$, will satisfy the KKT conditions of $Q_2(t)$, Eq. (D.28)-(D.31), and the KKT conditions of $Q_2(t^+)$, Eq. (D.35)-(D.38),

$$\nu^*_{n,t} = \hat{\nu}^*_{n,t}, \tag{D.61}$$

$$\mu^*_{n,t} = \hat{\mu}^*_{n,t}, \tag{D.62}$$

$$\alpha^*_{n,t} = \hat{\alpha}^*_{n,t}, \tag{D.63}$$

$$\beta^*_{n,t} = \hat{\beta}^*_{n,t}. \tag{D.64}$$

This means the energy production and expansion corresponding to the SPNE of the game induced by the mechanism are solutions of Problem (Socially Optimal - Longterm Uncertainty II).

Proof of Theorem 15 To prove that the mechanism is individually rational, we need to show that starting from each time t or t^+, $t \in T$, GenCo n's expected future utility due to the SPNE strategy of the game induced by the mechanism is greater than or equal to its non-continuation utility (which is 0).

First consider the expansion epoch t. GenCo n's expected utility at equilibrium starting from time t is determined by the solution of $\hat{Q}_2(t)$ as

$$\hat{V}^*_{n,t} = E_{\mathbf{w}_{t:T}|w_{1:t-1}}\left[\sum_{\tau\geq t}\beta^\tau\left(-C^e_{n,\tau}(\hat{e}^*_{n,\tau}, w_\tau) - C^x_{n,\tau}(\hat{x}^*_{n,\tau}, w_{\tau-1}) + \hat{p}^*_\tau \hat{e}^*_{n,\tau}\right)\Bigg| h_t\right]. \tag{D.65}$$

To show this is positive for all $t \in T$, consider the following surrogate optimization for GenCo n where the prices $p^*_\tau, \tau \geq t$ are fixed and equal to those corresponding to the SPNE strategies.

$$\begin{aligned}
\max_{\hat{x}'_{n,\tau},\hat{e}'_{n,\tau}} \quad & E_{\mathbf{w}_{t:T}|w_{1:t-1}}\left[\sum_{\tau\geq t}\beta^\tau\left(-C^e_{n,\tau}(\hat{e}'_{n,\tau}) - C^x_{n,\tau}(\hat{x}'_{n,\tau}, w_{\tau-1}) + \hat{p}^*_\tau \hat{e}'_{n,\tau}\right)\right] \\
s.t. \quad & 0 \leq \hat{x}'_{n,\tau} \leq \bar{x}_{n,\tau} \\
& 0 \leq \hat{e}'_{n,\tau} \leq \hat{x}_{n,t-1} + \sum_{\tau'\in\{t,t+1,\ldots,\tau\}} \hat{x}_{n,\tau'}
\end{aligned} \tag{D.66}$$

Call this problem $\hat{Q}^{sur}_2(t)$. Problem $\hat{Q}^{sur}_2(t)$ is a convex optimization satisfying the constraint qualification conditions. Note that because of Theorems (13) and (14), GenCo n's SPNE strategies, $\hat{\sigma}^*_{n,t}$ and $\hat{\gamma}^*_{n,t}$, are a solution of $\hat{Q}_2^{sur}(t)$, and $\hat{V}_{n,t}$ is equal to maximum value of $\hat{Q}_2^{sur}(t)$. Furthermore, note that $(0,0)$ is a feasible solution of problem $\hat{Q}_2^{sur}(t)$ and the value of $\hat{Q}_2^{sur}(t)$ at $(0,0)$ is equal to zero. Consequently, $\hat{V}_{n,t}$ is non-negative, therefore GenCo n will continue participating in the mechanism at time t.

Next, consider the generation at time t^+. The expected utility at time t^+ is determined by

$$\begin{aligned}
\hat{V}(n,t^+) &= -C^e_{n,\tau}(\hat{e}^*_{n,\tau}) + \hat{p}^*_\tau \hat{e}^*_{n,\tau} \\
&+ E_{\mathbf{w}_{t+1:T}|w_{1:t}}\left[\sum_{n\in N,\tau\geq t+1}\beta^\tau\left(-C^e_{n,\tau}(\hat{e}^*_{n,\tau}) - C^x_{n,\tau}(\hat{x}^*_{n,\tau}, w_{\tau-1}) + \hat{p}^*_\tau \hat{e}^*_{n,\tau}\right)\right].
\end{aligned} \tag{D.67}$$

By Theorem (4), $-C^e_{n,\tau}(\hat{e}^*_{n,\tau}) + \hat{p}^*_\tau \hat{e}^*_{n,\tau}$, which is GenCo n's maximum utility in the spot market at

t^+, is non-negative. In addition, the expected utility from time $t+1$ on is

$$E_{\mathbf{w}_{t+1:T}|w_{1:t}}\left[\sum_{n\in N,\tau\geq t+1}\beta^{\tau}\left(-C^{e}_{n,\tau}(\hat{e}^{*}_{n,\tau})-C^{x}_{n,\tau}(\hat{x}^{*}_{n,\tau},w_{\tau-1})+\hat{p}^{*}_{\tau}\hat{e}^{*}_{n,\tau}\right)\right], \tag{D.68}$$

and this term is non-negative by the first part of this proof. Therefore, $\hat{V}(n,t^+)$ is non-negative and GenCo n will continue participating in the mechanism at time t^+.

Proof of Theorem 16 Problem Q_2 is strictly convex and therefore it has a unique solution. Due to Theorem 14, at any SPNE of the game induced by the mechanism

$$\hat{e}^{*}_{n,t}=e^{*}_{n,t} \tag{D.69}$$

$$\Delta\hat{x}^{*}_{n,t}=\Delta x^{*}_{n,t} \tag{D.70}$$

and due to Theorem 1

$$\hat{p}^{*}_{n,t}=U'_t(\sum_{n\in N}e^{*}_{n,t}). \tag{D.71}$$

These equations uniquely determine the SPNE of the game induced by the mechanism.

Proof of Theorem 17 Theorem 14 shows that the NE of the game induced by the mechanism results in the same expansion and generation as the optimal solution of the corresponding social optimality problem, Problem (Socially Optimal - Longterm Uncertainty II). Therefore, to prove Theorem (17), we show that for Problem (Socially Optimal - Longterm Uncertainty II) and its optimal strategy, if $\Delta x^{*}_{n,t}(h_t)=\sigma^{*}_{n,t}(h_t)>0$ then there exists at least one history $h_\tau\in H_\tau,\quad \tau\geq t$ with $P(h_{t:\tau}|h_t)>0$ such that GenCo n is saturated, i.e.

$$\gamma^{*}_{n,\tau}(h_\tau)=e^{*}_{n,\tau}(h_t)=x^{*}_{n,\tau}. \tag{D.72}$$

We prove this by contradiction. Assume $\gamma^{*}_{n,\tau}(h_\tau)=e^{*}_{n,\tau}(h_\tau)<x^{*}_{n,\tau},\forall\tau,h_\tau\in H_\tau$ such that $P(h_{t:\tau}|h_t)>0$. We construct an alternative feasible strategy by adjusting the optimal strategy that can improve the social welfare by reducing the expansion strategy only at history h_t, $\Delta x^{*}_{n,t}(h_t)=\sigma^{*}_{n,t}(h_t)$, by small enough ε such that the constraints in Problem (Socially Optimal - Longterm Uncertainty II) are not violated (this is possible since by assumption the capacity is never saturated in the histories with positive probability). This way we reduce the expansion cost at time t but all the future costs for production and expansion and all the payments remain the same, thus the social welfare increases. This contradicts the fact that $(\gamma^{*}_{n,t},\sigma^{*}_{n,t},\quad n\in N,t\in T)$ is an optimal solution of Problem (Socially Optimal - Longterm Uncertainty II).