RISK-AVERSE MULTI-ARMED BANDITS AND GAME THEORY

BY

ALI YEKKEHKHANY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Professor Rakesh Nagi, Chair
Professor Bruce Hajek
Assistant Professor Ilan Shomorony
Professor Rayadurgam Srikant

ABSTRACT

The multi-armed bandit (MAB) and game theory literature is mainly focused on the expected cumulative reward and the expected payoffs in a game, respectively. In contrast, the rewards and the payoffs are often random variables whose expected values only capture a vague idea of the overall distribution. The focus of this dissertation is to study the fundamental limits of the existing bandits and game theory problems in a risk-averse framework and propose new ideas that address the shortcomings. The author believes that human beings are mostly risk-averse, so studying multi-armed bandits and game theory from the point of view of risk aversion, rather than expected reward/payoff, better captures reality. In this manner, a specific class of multi-armed bandits, called explore-then-commit bandits, and stochastic games are studied in this dissertation, which are based on the notion of Risk-Averse Best Action Decision with Incomplete Information (R-ABADI, Abadi is the maiden name of the author's mother). The goal of the classical multi-armed bandits is to exploit the arm with the maximum score defined as the expected value of the arm reward. Instead, we propose a new definition of score that is derived from the joint distribution of all arm rewards and captures the reward of an arm relative to those of all other arms. We use a similar idea for games and propose a risk-averse R-ABADI equilibrium in game theory that is possibly different from the Nash equilibrium. The payoff distributions are taken into account to derive the risk-averse equilibrium, while the expected payoffs are used to find the Nash equilibrium. The fundamental properties of games, e.g. pure and mixed risk-averse R-ABADI equilibrium and strict dominance, are studied in the new framework and the results are expanded to finite-time games. Furthermore, the stochastic congestion games are studied from a risk-averse perspective and three classes of equilibria are proposed for such games. It is shown by examples that the risk-averse behavior of travelers in a stochastic congestion game can improve

the price of anarchy in Pigou and Braess networks. Furthermore, the Braess paradox does not occur to the extent proposed originally when travelers are risk-averse.

We also study an online affinity scheduling problem with no prior knowledge of the task arrival rates and processing rates of different task types on different servers. We propose the Blind GB-PANDAS algorithm that utilizes an exploration-exploitation scheme to load balance incoming tasks on servers in an online fashion. We prove that Blind GB-PANDAS is throughput optimal, i.e. it stabilizes the system as long as the task arrival rates are inside the capacity region. The Blind GB-PANDAS algorithm is compared to FCFS, Max-Weight, and c-μ-rule algorithms in terms of average task completion time through simulations, where the same exploration-exploitation approach as Blind GB-PANDAS is used for Max-Weight and c-μ-rule. The extensive simulations show that the Blind GB-PANDAS algorithm conspicuously outperforms the three other algorithms at high loads.

To my mother, for her endless love and support.

ACKNOWLEDGMENTS

I want to thank my dissertation advisor, Professor Rakesh Nagi, for his guidance, brilliant ideas, and support. Professor Nagi has always been there to help me brainstorm through research and help me move forward. I also thank the committee members, Professor Bruce Hajek, Assistant Professor Ilan Shomorony, and Professor Rayadurgam Srikant, for their constructive comments. The professional publication editors of the ECE department, Jamie Hutchinson and Jan Progen, have been of great help. Last but not the least, I thank my mother, sister, and friends for their love and support.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

CB	CVaR Best-response
CVaR	Conditional Value at Risk Level
ETC	Explore-Then-Commit
FCFS	First-Come-First-Served
FTE	Finite-Time Exploitation
JSQ-MW	Join the Shortest Queue-MaxWeight
MAB	Multi-Armed Bandit
MB	Mean-variance Best-response
MV	Mean-Variance
OTE	One-Time Exploitation
PANDAS	Priority Algorithm for Near-Data Scheduling
R-ABADI	Risk-Averse Best Action Decision with Incomplete Information
RB	Risk-averse Best-response

Chapter 1

INTRODUCTION

1.1 Introduction and Contributions

The multi-armed bandit (MAB) and game theory literature is mainly focused on the expected cumulative reward and the expected payoffs in a game, respectively. In contrast, the rewards and the payoffs are often random variables whose expected values are vague representations of the overall distributions and do not capture the mean-variance trade-off that is associated with the risk of taking a specific action. The focus of this dissertation is to study the fundamental limits of the existing bandits and game theory problems in a risk-averse framework and propose new ideas to overcome the issues. The author believes that the human beings mostly behave in risk-averse manners, so studying multi-armed bandits and games from a risk-aversion point of view better captures reality. Risk-averse algorithms for MAB problems and risk-averse equilibria for (congestion) games are introduced in this dissertation, which are based on the notion of Risk-Averse Best Action Decision with Incomplete Information (R-ABADI[1]).

Multi-armed bandits have a wide range of applications as diverse as configuring web interfaces, paging and caching, routing in both wired and wireless networks, data structures, advertisement placement, dynamic pricing and online auction mechanisms, experiment design, and recommender systems, to name a few. A specific class of multi-armed bandits, called explore-then-commit (ETC) bandits, is studied in Chapter 2. This class of bandits is widely used in autonomous vehicles, clinical trial design, and investment companies. The objective in classical multi-armed bandit problems is to exploit the arm with the maximum expected reward. However, the expected reward does not capture the risk associated with arm rewards. As a result,

[1] Abadi is the maiden name of the author's mother.

if the objective of a player is not to maximize the cumulative reward, but to have a balanced reward in each and every play of the arms, or if the player only exploits one arm once after a pure exploration phase, then exploiting the arm with the maximum expected reward may no longer be desirable. The goal in explore-then-commit finite bandits is to identify the best arm after a pure experimentation phase to exploit it once or for a given finite number of times. In this setting, we observe that pulling the arm with the highest expected reward is not necessarily the most desirable objective for exploitation. Alternatively, we advocate the idea of risk aversion where the objective is to compete against the arm with the best risk-return trade-off. We propose a class of hyper-parameter-free risk-averse algorithms, called OTE/FTE-MAB (One/Finite-Time Exploitation Multi-Armed Bandit), whose objectives are to select the arm that will most likely provide the greatest reward in a single or finite-time exploitation. To analyze these algorithms, we define a new notion of finite-time exploitation regret for our setting of interest. We provide an upper bound of order $\ln\left(\frac{1}{\epsilon_r}\right)$ for the minimum number of experiments that should be done to guarantee upper bound ϵ_r for regret. In contrast to the existing risk-averse bandit algorithms, our proposed algorithms do not rely on hyper-parameters, resulting in a more robust behavior in practice. In the case that pulling an arm in the exploration phase has a cost, a trade-off between cost and regret emerges. We propose the c-OTE-MAB algorithm for two-armed bandits that addresses the cost-regret trade-off by minimizing a linear combination of cost and regret, using a hyper-parameter, that is called *cost-regret function*. This algorithm estimates the optimal number of explorations whose cost-regret value approaches the minimum value of the cost-regret function at the rate $\frac{1}{\sqrt{n_e}}$ with an associated confidence level, where n_e is the number of explorations of each arm.

The risk-averse R-ABADI equilibrium for stochastic games is studied in Chapter 3. The term *rational* has become synonymous with maximizing expected payoff in the definition of the best response in the Nash setting. In this chapter, we consider stochastic games in which players engage only once, or at most a limited number of times. In such games, it may not be rational for players to maximize their expected payoff as they cannot wait for the law of large numbers to take effect. We instead introduce probability statements on the best response by defining a new notion of a risk-averse best response that takes the payoff distributions into account. This results

in a risk-averse R-ABADI equilibrium in which players choose to play the strategy that maximizes the probability of their being rewarded the most in a single round of the game rather than maximizing the expected received reward, subject to the actions of other players. Note that the psychology of risk-averse players is such that they do not often take the expected payoffs into account, but consider the payoff distributions instead. A strategy with high expected payoff may be less likely to have a higher payoff than another strategy with lower expected payoff, which is shown in an illustrative example in this chapter. The Nash equilibrium is such that no player has any incentive to deviate from his/her strategy since all his/her strategies have the same expected payoff given the other players' strategies. In contrast, the risk-averse R-ABADI equilibrium makes each player indifferent to his/her choice of strategies by giving all strategies the same probability of rewarding more than or equal to all the other strategies, given the other players' strategies. We show that the risk-averse equilibrium based on the mentioned probability statement can be found by realizing the Nash equilibrium of a new game whose payoffs are derived from the probability distributions of the payoffs of the original game.

Stochastic congestion games are studied in Chapter 4. The fast-growing market of autonomous vehicles, unmanned aerial vehicles, and fleets in general necessitates the design of smart and automatic navigation systems considering the stochastic latency along different paths in the traffic network. The longstanding shortest path problem in a deterministic network, whose counterpart in a congestion game setting is Wardrop equilibrium, has been studied extensively, but it is well known that finding the notion of an optimal path is challenging in a traffic network with stochastic arc delays. In order to address this issue, several researchers have attempted to take the uncertainty in travel times into account when defining the notion of best response in a stochastic congestion game by considering a safety margin to arrive on time, the probability of being late/on time, or adding mean travel time and an additional component related to the variance of travel time. However, simplifying assumptions are made in these works such as considering the arc delay distributions to be independent of their loads or adding independent and identically distributed errors to nominal delays of arcs neglecting their differences. In this chapter, we propose three classes of risk-averse equilibria for an atomic stochastic congestion game in its general case where the

arc delay distributions are load dependent and not necessarily independent of each other. The three classed are R-ABADI equilibrium, mean-variance equilibrium (MVE), and conditional value at risk level α equilibrium ($\text{CVaR}_\alpha\text{E}$) whose notions of risk-averse best responses are based on maximizing the probability of taking the shortest path, minimizing a linear combination of mean and variance of path delay, and minimizing the expected delay at a specified risky quantile of the delay distributions, respectively. We prove that for any finite stochastic atomic congestion game, the risk-averse, mean-variance, and CVaR_α equilibria exist. We show that for risk-averse travelers, the Braess paradox may not occur to the extent presented originally since players do not necessarily travel along the shortest path in expectation, but they take the uncertainty of travel time into consideration as well. Although the focus of this work is not on deriving bounds on price of anarchy, we show through some examples that the price of anarchy/social delay can be improved when players are risk-averse and travel according to one of the three classes of risk-averse equilibria rather than when travelers are risk-neutral/selfish and travel according to the Wardrop equilibrium.

In Chapter 5, an exploration-exploitation scheme is used for load balancing with incomplete knowledge of the processing rates of tasks on servers. Dynamic affinity load balancing of multi-type tasks on multi-skilled servers, when the service rate of each task type on each of the servers is known and can possibly be different from the others, has been an open problem for over three decades. The goal is to do task assignment on servers in a real-time manner so that the system becomes stable, which means that the queue lengths do not diverge to infinity in steady state (throughput optimality), and the mean task completion time is minimized (delay optimality). The fluid model planning, Max-Weight, and c-μ-rule algorithms have theoretical guarantees on optimality in some aspects for the affinity problem, but they consider a complicated queueing structure and require either the task arrival rates, the service rates of tasks on servers, or both. In many real-world applications, both task arrival rates and service rates of different task types on different servers are unknown. To tackle this issue, we propose the Blind GB-PANDAS algorithm which is completely blind to task arrival rates and service rates. Blind GB-PANDAS uses an exploration-exploitation approach for load balancing. We prove that Blind GB-PANDAS is throughput optimal under arbitrary and unknown distributions for service times of different

task types on different servers and unknown task arrival rates. Blind GB-PANDAS aims to route an incoming task to the server with the minimum weighted-workload, but since the service rates are unknown, such routing of incoming tasks is not guaranteed, making the throughput optimality analysis more complicated than in the case where service rates are known. The extensive experimental results reveal that Blind GB-PANDAS significantly outperforms existing methods in terms of mean task completion time at high loads.

1.2 Related Work

We present the related work on multi-armed bandits, risk-averse stochastic games, stochastic congestion games, and affinity load balancing in the following subsections.

1.2.1 Budgeted Explore-Then-Commit Bandits and Risk-Averse Algorithms

Explore-then-commit bandit is a class of multi-armed bandit problems that has two consecutive phases called *exploration* (experimentation) and *commitment* [1, 2]. The decision-maker can arbitrarily explore each arm in the experimentation phase; however, he/she needs to commit to one selected arm in the commitment phase. There are few studies on explore-then-commit bandits in the literature that are summarized below. Bui et al. [1] studied the optimal number of explorations when cost is incurred in both phases. Liau et al. [3] designed an explore-then-commit algorithm for the case where there is a limited space to record the arm reward statistics. Perchet et al. [4] studied explore-then-commit policy under the assumption that the employed policy must split explorations into a number of batches. To the best of our knowledge, there is no cost-based study that considers risk-aversion in an explore-then-commit bandit with finite exploitations, which is the focus of this work. In the following, a review on bandits is presented that considers both arm rewards and the exploration-exploitation cost. The review is followed by an overview of risk-averse bandits.

A class of multi-armed bandit problems is associated with budget constraints where a player receives a reward with a cost by pulling an arm. Two types of MAB with budget constraint have been mainly studied [5]. First, pulling an arm is associated with a cost or constrained by a budget in both exploration and exploitation phases. Second, pulling an arm has a cost constrained by a budget only in the exploration phase. This type of MAB with budget constraint is called *pure exploration* or *best arm identification* [6]. There are several studies of the first type. Tran-Thanh et al. [7] proposed the ϵ-first algorithm for MAB with a limited budget imposed on pulling arms, where pulling each arm has a different fixed cost. In the ϵ-first algorithm, ϵ of the budget is used for the exploration phase and the remaining budget is used for the exploitation phase. The regret bound of the ϵ-first algorithm is $\mathcal{O}\left(B^{\frac{2}{3}}\right)$, where B is the budget value. The drawbacks of this algorithm are the polynomial regret bound and that a large ϵ assures a more accurate exploration but with an ineffective exploitation, and vice versa. In order to resolve these issues, Tran-Thanh et al. [5] used a knapsack setting and improved the regret bound from $\mathcal{O}\left(B^{\frac{2}{3}}\right)$ to $\mathcal{O}\left(\ln B\right)$. In another study, Ding et al. [6] considered the cost of pulling arms as a discrete random variable rather than a fixed cost and proposed the UCB-BV1 and UCB-BV2 algorithms. In the UCB-BV1 algorithm, the lower bound of the expected costs needs to be known, but the UCB-BV2 algorithm estimates this lower bound. The regret bound for both these algorithms is proven to be $\mathcal{O}\left(\ln B\right)$. Xia et al. [8] studied the limited budget setting in both multi-armed bandits and linear bandits with continuous random costs. They proposed the Budget-UCB and Budget-CB algorithms for MAB and linear bandit with distribution-dependent regret bound $\mathcal{O}\left(\ln B\right)$ and polylog(B), respectively. Additionally, Xia et al. [8] studied the limited budget MAB with multiple plays, where the player pulls multiple arms in each round. They proposed the MP-BMAB algorithm that uses a multiple ratio confidence bound to determine the best arms to play with a sublinear regret. Xia et al. [9] applied Thompson sampling to the limited budget MAB problem with random cost associated for pulling an arm and proposed the BTS algorithm that has a distribution-dependent regret bound of $\mathcal{O}\left(\ln B\right)$. In another work, Badanidiyuru et al. [10] studied the MAB problem with multiple budget constraints where the budget consumption of pulling an arm is a random multi-dimensional vector. They used the knapsacks model to address this

problem and proposed the PD-BwK algorithm with a sublinear regret. The above setting is called *bandits with knapsacks* and is extended in conceptual bandits by [11] and [12]. In the second type, best arm identification, there are several studies. Bubeck et al. [13] studied a case where a player explores arms with a limited budget without concern about the received rewards in order to identify the best arm after the pure exploration phase. To evaluate the best identified arm, they defined simple regret as the difference between the maximal expected reward and the expected reward of the best identified arm. They find upper bounds for the simple regret for two cases. In the first case, arms are played uniformly in the pure exploration phase and the best identified arm is the empirical best arm. In the second case, a UCB-based exploration is performed and the best identified arm is the most played arm. Audibert and Bubeck [14] defined the probability of selecting a suboptimal arm as regret for the pure exploration setting. They found upper bounds on the regret for both UCB exploration and their own proposed SR algorithm. Gabillon et al. [15] studied the best arm identification for each of the bandits in a multi-bandit multi-armed setting. They defined regret as the maximum error among all bandits, where error is defined as the probability of selecting a suboptimal arm. They proposed the GapE and GapE-V algorithms for exploration and obtained upper bounds on their regret. The GapE algorithm is UCB-based which takes into account the gap between the expected rewards of the optimal arm and the best identified arm. The GapE-V algorithm is also UCB-based and not only uses the gap but also considers the estimated reward variances.

There are several criteria to measure and to model risk in the risk-averse multi-armed bandit problem. One of the common risk measurements is the mean-variance paradigm [16]. The two algorithms MV-LCB and ExpExp proposed by Sani et al. [17] are based on the mean-variance concept. They define the mean-variance of an arm with mean μ and variance σ^2 as $\text{MV}= \sigma^2 - \rho \cdot \mu$, where $\rho \geq 0$ is the absolute risk tolerance coefficient. In an infinite horizon multi-armed bandit problem, MV-LCB plays the arm with minimum lower confidence bound for estimation of MV. In a best-arm identification setting, the ExpExp algorithm explores each of the arms for the same number of times and selects the arm with minimum estimated MV. This approach is followed by numerous researchers in risk-averse multi-armed bandit problems [18–21]. Another way of considering risk in multi-armed bandit

problems is to use conditional value at risk level α, CVaR_α, where it is the expected policy return in a specified quantile. CVaR_α is utilized by Galichet et al. [22] in risk-aware multi-armed bandit problems. They presented the Multi-Armed Risk-Aware Bandit (MaRaB) algorithm aiming to select the arm with the maximum conditional value at risk level α, CVaR_α. Formally, let $0 < \alpha < 1$ be the target quantile level and v_α defined as $\mathbb{P}(R < v_\alpha) = \alpha$ be the associated quantile value, where R is the arm reward. The conditional value at risk α is then defined as $\text{CVaR}_\alpha = \mathbb{E}[R|R < v_\alpha]$. CVaR_α is also followed by researchers in multi-armed bandit problems [18, 23–26]. The performances of both MV and CVaR are highly dependent on different single scalar hyper-parameters, and selecting an inappropriate hyper-parameter might degrade the performance substantially. The negative impact of hyper-parameter mismatch is studied in Section 2.4.

1.2.2 Risk-Averse Equilibrium for Games

Since the seminal works of von Neumann [27], von Neumann and Morgenstern [28], and Nash [29], expected utility has emerged as the dominant objective value within game theory as each player attempts to maximize his/her expected utility given the actions of other players. This concept was extended naturally into games of incomplete information (Bayesian games) by Harsanyi [30], as players can still maximize their expected utility given a distribution from which the game will be drawn. These games have received a great deal of attention as they more accurately model real-world situations where not all parameters are known precisely, with later works such as Wiseman [31] addressing how players sequentially refine their equilibria as they learn the distributions and the more recent work of Mertikopoulos and Zhou [32] addressing how players learn their payoffs with continuous action sets. Another recent work by Sugaya and Yamamoto [33] considers the more specific question of how firms in a duopoly should play when the payoff distributions are based on the market state, a random variable with possibly unknown distribution.

Despite all the work that has gone into expected utility as the objective value players wish to maximize, it is still questionable whether this is a good assumption [34, 35]. Several papers have focused on adding a degree of risk

aversion to player's utility functions in specific games, with Angelidakis et al. [36] and Bell and Cassir [37] analyzing variations of this in congestion games, and Yamazaki [38] doing so in rent-seeking games, Harrington [39] doing so in bargaining games. Additionally Goeree et al. [40] present an empirical study of risk-aversion in the matching-pennies game, where they observe marked deviations from Nash behavior (expected utility maximization) as the payoffs/costs become larger. This is consistent with the concept of prospect theory based on empirical observations across several experiments in which the subjects deviate from actions which would maximize their expected utility. Kahneman and Tversky [41] formulated the idea of prospect theory, which states that consumers are naturally risk-averse when addressing situations with potential gains and naturally risk-seeking when facing situations with potential losses. Prospect theory has since been widely studied, with an extension of the original paper provided in [42] to address more general payoff/cost functions. Levy [43] provides a good overview of classical prospect theory, particularly from a political perspective. Unsurprisingly, prospect theory has received a great deal of attention in financial studies [44, 45], with Barberis et al. [46] using it for asset pricing. Prospect theory is not without its critics; e.g., List [47] posits that the results of the studies on prospect theory are due to inexperienced consumers, and designs an experiment to show these behaviors disappear with experience. However, experienced consumers are by definition consumers who engage in similar trials multiple times, which means that for these consumers expected utility *is* an appropriate metric. As we are explicitly interested in games which will be played at most a small number of times, we do not need to be concerned with this effect.

1.2.3 Stochastic Congestion Games

In this section, the literature on navigation for both deterministic and stochastic networks is presented first, then the literature on deterministic and stochastic congestion games is discussed in detail. The main focus of the literature review is to motivate the necessity of risk-averse algorithms for navigation and congestion games in a stochastic setting.

The problem of finding the shortest path in a transportation or telecom-

munication traffic network is one of the main parts of the in-vehicle navigation systems. This problem has been studied well in deterministic networks resulting in many efficient algorithms, e.g., the algorithms developed by Bellman [48], Dijkstra [49], and Dreyfus [50], also see [51–60]. Although finding the shortest path problem is well understood in deterministic networks, the definition of an optimal path and how to identify such a path is more challenging in the stochastic version of the problem. There have been multiple approaches to define the optimal path in stochastic networks as summarized below. The least expected travel time is studied by Loui [61] and is equivalent to the deterministic case from the computational point of view. The path with the least expected time may be sub-optimal for risk-averse travelers due to its high variability and uncertainty; as the result, the probability distributions of link travel times need to be considered explicitly to find the most reliable path. In this manner, Frank [62] proposed the optimal path to be the one that maximizes the probability of realizing a travel time that less than a threshold, Sigal et al. [63] proposed the optimal path to be the one that maximizes the probability of realizing the shortest time, and Chen and Ji [64] proposed the optimal path to be the one with minimum travel time budget required to meet a travel time reliability constraint. For more variants of the mentioned algorithms, refer to [65–81].

In the context of route selection in a fleet of vehicles, a game emerges between all travelers where the action of each traveler affects the travel time of the other travelers, which creates a competitive situation forcing travelers to strategize their decisions. In a deterministic network, the mentioned game is formalized by Wardrop and Whitehead [82], von Neumann [83], von Neumann and Morgenstern [84], and Nash et al. [85]. However, it is not realistic to consider the link delays to be known prior to making a decision due to external factors that make the travel times uncertain. In order to put this in perspective, several approaches have been adopted by researchers to capture the stochastic behavior of the traffic networks. For example, Harsanyi [86, 87] proposed Bayesian games that consider the incomplete information of payoffs, Ordóñez and Stier-Moses [88] modeled the risk-averse behavior of travelers by padding the expected travel time along paths with a safety margin, Watling [89] proposed an equilibrium based on the optimality measure of minimizing the probability of being late or maximizing the probability of being on time, Szeto et al. [90] associated a cost with the travel

time uncertainty based on travelers' risk-averse behavior, Chen and Zhou [91] proposed an equilibrium based on the optimality measure of minimizing the conditional expectation of travel time beyond a travel time budget, and Bell and Cassir [92] proposed to play out all possible scenarios before making a choice. For more details in the context of traffic networks, we refer readers to [93–105].

1.2.4 Affinity Scheduling Algorithms

There is a huge body of work on a specific class of affinity problems with applications in scheduling for data centers considering data locality, which can be divided into two main categories: (1) Heuristic scheduling algorithms with no theoretical guarantees on throughput or delay optimality, see e.g. [106–116]. Although some of these heuristic algorithms are being used in real applications, simple facts about their optimality are not investigated. Among these algorithms, the Fair Scheduler is the de facto standard in Hadoop [108]. Other than map task scheduling for map-intensive jobs, heuristic algorithms like [117–119] study the joint scheduling of map and reduce tasks. (2) Algorithms that theoretically guarantee throughput or delay optimality or both [120–148]. The works by Harrison [120], Harrison and Lopez [121], and Bell and Williams [122, 123] on affinity scheduling not only require the knowledge of mean arrival rates of all task types, but also consider one queue per task type. In a data center, if a task is replicated on three servers, the number of task types can be in the cubic order of number of servers, which causes unnecessary and intolerable complexity of the system. The MaxWeight algorithm (the generalized cμ-rule) by Stolyar and Mandelbaum [124, 127] does not require the arrival rates, but still needs one queue per task type. The JSQ-MaxWeight algorithm by Wang et al. [132] solves the per-task-type problem for a system with two levels of data locality. The JSQ-MaxWeight algorithm is throughput optimal, but it is delay optimal for a special traffic scenario. The priority algorithm for near data scheduling [133] is both throughput and heavy-traffic optimal for systems with two locality levels. The weighted-workload routing and priority scheduling algorithm [134] for systems with three locality levels is shown to be throughput optimal and delay optimal in a larger region of the capacity region compared

to the JSQ-MW algorithm with a further assumption on the processing rates.

A related direction of work on scheduling for data centers with multi-level data locality, which is a direct application of affinity scheduling, is to efficiently do data replication on servers in MapReduce framework to increase availability. Increasing the availability is translated to increasing service rates in the context of this dissertation which enlarges the capacity region and reduces the mean task completion time. For more information on data replication algorithms refer to Google File System [149], Hadoop Distributed File System [106], Scarlett [150], and Dare [151]. Data replication algorithms are complementary and orthogonal to the throughput and delay optimality that is studied in this dissertation.

In addition to data-locality, fairness is another concern in task scheduling which actually conflicts with delay optimality. A delay optimal load balancing algorithm can cooperate with fair scheduling strategies, though, by compromising on delay optimality to achieve partial fairness. For more details on fair scheduling [152–154], see Zaharia et al. [108], Isard et al. [107], and the references therein. A different line of work studies the performance of load balancing algorithms under uncertainty of system parameters. It is desirable to design algorithms that are robust to the changes in task arrival loads, change of service rates, and other factors. Some robust policies are studied in [155–159].

1.2.5 Applications of Affinity Scheduling in MapReduce Framework

In large scale data-intensive applications like the healthcare industry, ad placement, online social networks, large-scale data mining, machine learning, search engines, and web indexing, the de facto standard is the MapReduce framework. MapReduce framework is implemented on tens of thousands of machines (servers) in systems like Google's MapReduce [160], Hadoop [106], and Dryad [161] as well as grid-computing environments [107]. Such vast investments do require improvements in the performance of MapReduce, which gives them new opportunities to optimize and develop their products faster [150]. In MapReduce framework, a large data-set is split into small data chunks (typically 64 or 128 MB) and each one is saved on a number of

machines (three machines by default) which are chosen uniformly at random. A request for processing the large data-set, called a job, consists mainly of two phases, map and reduce. The map tasks read their corresponding data chunks which are distributed across machines and output intermediary key-value results. The reduce tasks aggregate the intermediary results produced by map tasks to generate the job's final result.

In MapReduce framework, a master node (centralized scheduler) assigns map and reduce tasks to slaves (servers) in response to heartbeats received from slaves. Since jobs are either map-intensive or only require map tasks [162, 163], and since map tasks read a large volume of data, we only focus on map task scheduling as an immediate application of our load balancing algorithm. Local servers of a map task refer to those servers having the data associated with the map task. Local servers process map tasks faster, so the map tasks are preferred to be co-located with their data chunks or at least be assigned to machines that are close to map tasks' data, which is commonly referred to as near-data scheduling or scheduling with data locality.

In contrast to the improvements in the speed of data center networks, there is still a huge difference between accessing data locally and fetching it from another server [164, 165]. Hence, improving data locality increases system throughput, alleviates network congestion due to less data transmission, and enhances users' satisfaction due to less delay in receiving their job's response. There are two main approaches to increase data locality: (1) Employing data replication algorithms to determine the number of data chunk replicas and where to place them (instead of choosing a fixed number of machines uniformly at random, which is done in Google File System [149] and Hadoop Distributed File System [106]). For more details see the algorithms Scarlett [150] and Dare [151]. (2) Scheduling map tasks on or close to local servers in a way to keep balance between data-locality and load-balancing (assigning all tasks to their local machines can lead to hotspots on servers with popular data). These two methods are complementary and orthogonal to each other. The focus of this dissertation is on the second method.

Chapter 2

A COST-BASED ANALYSIS FOR RISK-AVERSE EXPLORE-THEN-COMMIT FINITE-TIME BANDITS

One of the classes of decision making models is the multi-armed bandit (MAB) framework where decision makers learn the model of different arms that are unknown and actions do not change the state of arms [166]. The MAB problem was originally proposed by Robbins [167], and has a wide range of applications in finance [168,169], communication and networks [170], health-care [171], autonomous vehicles [172], dynamic spectrum access systems [173], and energy management [22, 174, 175] to name but a few. In the classical MAB problem, the decision-maker sequentially selects an arm (action) with an unknown reward distribution out of K arms. The noisy reward of the selected arm is revealed and the values of other arms remain unknown. At each step, the decision-maker encounters a dilemma between exploitation of the best identified arm versus exploration of alternative arms. The goal of the classical model of multi-armed bandit is to maximize the expected cumulative reward over a time horizon.

In this chapter, the focus is on a setting where a player is allowed to explore different arms in the exploration (or experimentation, used interchangeably) phase before committing to the best identified arm for exploitation in one or a given finite number of times. A preliminary treatment of this problem has been introduced in [176]. Besides, pulling an arm in the exploration phase can incur a cost. This setting of interest is motivated by several application domains such as personalized health-care and one-time investment. In such applications, exploitation is costly and/or it is infeasible to exploit for a large number of times, but arms can be tested by simulation and/or based on the historical data for multiple times with a relatively small cost [1]. The big step in personalized health-care is to provide an individual patient with his/her disease risk profile based on his/her electronic medical record

Portions of this chapter were previously published in Yekkehkhany et al. [176] and is used here with permission.

and personalized assessments [177, 178]. The different treatments (arms) are evaluated for a person by simulation or mice trials for many times with a low cost, but one personalized treatment is exploited once for a patient in the end [179, 180]. Another example of one-time exploitation is one-time investment where an investor chooses a factory out of multiple ones. Based on experimentation on historical data, he/she selects a factory to invest in once. The common theme in both above examples is to identify the best arm for one-time exploitation after an experimentation phase of pure exploration. This setting falls in the class of MAB problems called *explore-then-commit* [1, 2].

Note that although pulling the arm with the maximum expected reward is desirable in the settings with infinite exploitations, it is not necessarily the best objective in the explore-then-commit setting with a single or finite exploitations. In such scenarios, players not only aim to achieve the maximum expected cumulative reward, but they also want to minimize the uncertainty such as risk in the outcome [18]. These approaches are known as *risk-averse* MAB. We advocate a risk-averse approach in which the objective is to *select an arm that is most probable to reward the most*. The previous works [1,2,17,22,181,182] on explore-then-commit bandits, to the best of our knowledge, try to identify the arm with *an optimal risk-return criterion* on an expectation sense up to a hyper-parameter. The choice of hyper-parameter is tricky which will be further elaborated by an illustrative example in Section 2.1. We propose a class of hyper-parameter-free risk-averse algorithms, called OTE/FTE-MAB, for explore-then-commit bandits with finite-time exploitations. We define a new notion of finite-time exploitation regret for our setting of interest and obtain an upper bound for the minimum number of experiments that should be done to guarantee an upper bound for regret that are elaborated in Section 2.2.

In the mentioned single or finite exploitation explore-then-commit bandit applications, although the exploration cost of arms is relatively small, a trade-off between cost and regret emerges at large numbers of explorations. Increasing the number of explorations decreases regret but increases cost and vice versa. In order to capture this issue, we formalize this trade-off for a two-armed bandit problem and propose an algorithm to determine an estimation of the optimal number of explorations. The cost-regret trade-off is studied in details in Section 2.3.

The contributions of this chapter are summarized below. We propose a class of hyper-parameter-free risk-averse algorithms (called `OTE/FTE-MAB`) for explore-then-commit bandits with finite-time exploitations. The goal of the algorithms is to select the arm that is most probable to give the player the highest reward. To analyze the algorithms, we define a new notion of finite-time exploitation regret for our setting of interest. We provide concrete mathematical support to obtain an upper bound of order $\ln(\frac{1}{\epsilon_r})$ for the minimum number of experiments that should be done to guarantee upper bound ϵ_r for regret. As a salient feature, the `OTE/FTE-MAB` algorithm is hyper-parameter-free, so it is not prone to errors due to the hyper-parameter mismatch. We further study the case where the cost of pulling arms in the exploration phase is not negligible, and as a result, a trade-off between cost and regret should be considered. We propose the `c-OTE-MAB` algorithm for a two-armed bandit that addresses this trade-off by minimizing a linear combination of cost and regret, using a hyper-parameter, that is called cost-regret function. This algorithm determines an estimation of the optimal number of explorations whose cost-regret value approaches the minimum value of the cost-regret function at the rate $\frac{1}{\sqrt{n_e}}$ with an associated confidence level, where n_e is the number of explorations of each arm. The c-OTE-MAB algorithm is designed for one-time exploitation that can be extended to finite-time exploitations.

The rest of the chapter is organized as follows. Subsection 1.2.1 discusses related work. In Section 2.1, the one/finite-time exploitation multi-armed bandit problem after an experimentation phase is formally described. We define a new notion of one/finite-time exploitation regret for our problem setup. An example is provided clarifying the motivation of our work. In Section 2.2, we propose the OTE-MAB and FTE-MAB algorithms, and find an upper bound for the minimum number of pure explorations needed to guarantee an upper bound for regret. In Section 2.3, we propose the c-OTE-MAB algorithm that determines an estimation of the optimal number of explorations for a two-armed bandit problem, where exploring arms is associated with a cost. In Section 2.4, we evaluate the OTE-MAB algorithm versus risk-averse baselines and compare the minimum number of experiments needed to guarantee an upper bound on regret for both the OTE-MAB and FTE-MAB algorithms. Additionally, we show by an example that the expected value of the estimated optimal number of explorations derived from the c-OTE-MAB

algorithm converges to the optimal value of the number of explorations. For conclusion of the chapter and a discussion of opportunities for future work, refer to Chapter 6.

2.1 Formulation of Explore-Then-Commit Finite Bandits

Consider arms $\mathcal{K} = \{1, 2, \dots, K\}$ whose rewards are random variables $R_1, R_2, \dots, R_K$ that have an unknown joint distribution function $f_{1,2,\dots,K}\big(u_1, u_2, \dots, u_K\big)$ with marginal distribution functions $f_1(u), f_2(u), \dots, f_K(u)$ with unknown finite expected values $\mu_1, \mu_2, \dots, \mu_K$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2$, respectively. Note that a bandit with independent arms is a specific case of a bandit with dependent arms since the joint distribution of arms in the former case is given by $f_{1,2,\dots,K}(u_1, u_2, \dots, u_K) = f_1(u_1) \times f_2(u_2) \times \cdots \times f_K(u_K)$. The goal is to identify the best arm at the end of an experimentation phase that is followed by an exploitation phase, where the best arm is exploited for a given number of times, $M < \infty$. In the experimentation phase, all arms are sampled together for N independent times. Denote the observed reward of arm $k \in \mathcal{K}$ at sample $n \in \{1, 2, \dots, N\}$ of experimentation by $r_{k,n}$. The uniform exploration of all arms for the same number of times is a common practice in bandit problems with pure exploration [13, 17, 183]. Note that if arms are independent, rewards of different arms can be sampled independently from each other. In Section 2.2, the cost of exploration is not considered, but in Section 2.3, a two-armed bandit is studied where the cost of pulling the two arms for n times is formulated by $C(n)$ and $\lim_{n\to\infty} = \infty$.

Let $R_k^M = X_k^1 + X_k^2 + \cdots + X_k^M$, for $k \in \mathcal{K}$, where $(X_1^m, X_2^m, \dots, X_K^m)$ for $m \in \{1, 2, \dots, M\}$ are independent and identically distributed multivariate random variables and $(X_1^1, X_2^1, \dots, X_K^1) \sim f_{1,2,\dots,K}$. The optimal arm for M exploitations in the sense that maximizes the hyper-parameter-free probability of receiving the highest reward is

$$k^* = \arg\max_k \mathbb{P}(R_k^M \geq \boldsymbol{R}_{-k}^M), \tag{2.1}$$

where $\boldsymbol{R}_{-k}^M = \{R_1^M, R_2^M, \dots, R_{k-1}^M, R_{k+1}^M, \dots, R_K^M\}$ and R_k^M being greater than or equal to a vector means that it is greater than or equal to all elements

of the vector. The mentioned measure of optimality is called Risk-Averse Best Action Decision with Incomplete Information (R-ABADI). Let $p_k^M = \mathbb{P}(R_k^M \geq \boldsymbol{R}_{-k}^M)$ be the score of arm k. Given the above preliminaries, the finite-time exploitation regret is defined below.

Definition 1. *The finite-time exploitation regret, $r_M(\Delta p)$, is defined to be the probability that the score of the selected arm $\hat{k}$, where $\hat{k}$ is a random variable, deviates from the score of the optimal arm by a tolerance threshold $0 < \Delta p < 1$; i.e.,*

$$r_M(\Delta p) = \mathbb{P}\left(p_{k^*}^M - p_{\hat{k}}^M \geq \Delta p\right). \tag{2.2}$$

Regret can also be defined hyper-parameter-free as $r_M = \mathbb{P}\left(\hat{k} \neq k^*\right)$, in which case the theoretical results in Section 2.2 become distribution-dependent, which is discussed in detail in that section. Note that the above definitions of regret and arm optimality are different from the commonly used regret and optimality criteria in bandit problems. In the following, an example is presented that motivates the definition of the new notion of regret as well as the new optimality criteria for the finite-time exploitation setting.

2.1.1 Illustrative Example

As mentioned in the Introduction, although the arm with the highest expected reward is the optimal arm for utilization in infinite number of exploitations, it is not necessarily the one that is most probable to have the highest reward in a single or some finite number of exploitations. In the following example, two arms are considered such that $\mu_2 > \mu_1$, but it is more probable that a one-time exploitation of the first arm rewards us more than a one-time exploitation of the second arm. Hence, arm $\arg\max_k \mu_k$ is not necessarily the ideal arm for one-time exploitation let alone the arm with the maximum empirical mean, i.e. $\arg\max_k \frac{\sum_{n=1}^N r_{k,n}}{n}$.

Example 1. *Consider two arms with the following independent reward distributions:*

$$f_1(u) = \alpha e^{-2(u-3)^2} \cdot \mathbb{1}\{0 \leq u \leq 10\}$$
$$f_2(u) = \beta\left(3e^{-8(u-1)^2} + 2e^{-8(u-8)^2}\right) \cdot \mathbb{1}\{0 \leq u \leq 10\},$$

where α and β are constants for which each of the two distributions integrate to one and $\mathbb{1}\{.\}$ is the indicator function.

In Example 1, although the second arm has a larger mean than the first one, $\mu_2 \approx 3.8$ and $\mu_1 \approx 3$, the variance of reward received from the second arm is larger than that from the first one, which increases the risk of choosing the second arm for a one-time exploitation application. In fact, the first arm with lower mean is more probable to reward us more than the second arm since $\mathbb{P}(R_1 \geq R_2) \approx 0.6 > 0.5$. In general, a larger variance for the received reward is against the principle of risk-aversion where the objective is to keep a balance in the trade-off between the expected return and risk of an action [17]. Mean-variance is an existing approach to tackle this scenario. However, it has some drawbacks that are explained in detail in the following.

The mean-variance (MV) of arm k is defined as $\sigma_k^2 - \rho \cdot \mu_k$ that depends on the hyper-parameter $\rho \geq 0$, which is the absolute risk tolerance coefficient. The arm with the minimum MV value is defined to be optimal in this framework. The trade-off on ρ is that if it is set to zero, the arm with the minimum variance is selected. On the other hand, if ρ goes to infinity, the arm with the maximum expected reward is selected, which is the same as classical multi-armed bandit approach. Although the behavior of mean-variance trade-off is known for marginal values of ρ, it is not obvious what value of the hyper-parameter ρ keeps a desirable balance between return and risk. The choice of this hyper-parameter can be tricky and as will be shown in Section 2.4; an inappropriate choice can increase the regret dramatically. As a simple example, consider two arms with unknown parameters $\mu_1 = 10, \sigma_1^2 = 10, \mu_2 = 1, \sigma_2^2 = 1$, and $\mathbb{P}(R_1 > R_2) = 1$. The mean-variance trade-off is formalized as $\hat{\sigma}_k^2 - \rho\hat{\mu}_k$, where $\hat{\sigma}_k^2$ and $\hat{\mu}_k$ are empirical estimates of variance and mean of each arm. Note that the empirical means and variances converge to true values, so the second arm that is performing worse with probability one is selected in limit if $\rho < 1$. The mean-variance framework aims at keeping a balance on choosing an arm with low variance and high expected reward. However, the limitation of this method is that high variance is not necessarily against the player. This fact is presented in another example depicted in Figure 2.1, where the blue arm, R_1, rewards more than the red arm, R_2, with probability one, so a logical player would choose the first arm to play. However, the mean-variance framework would choose

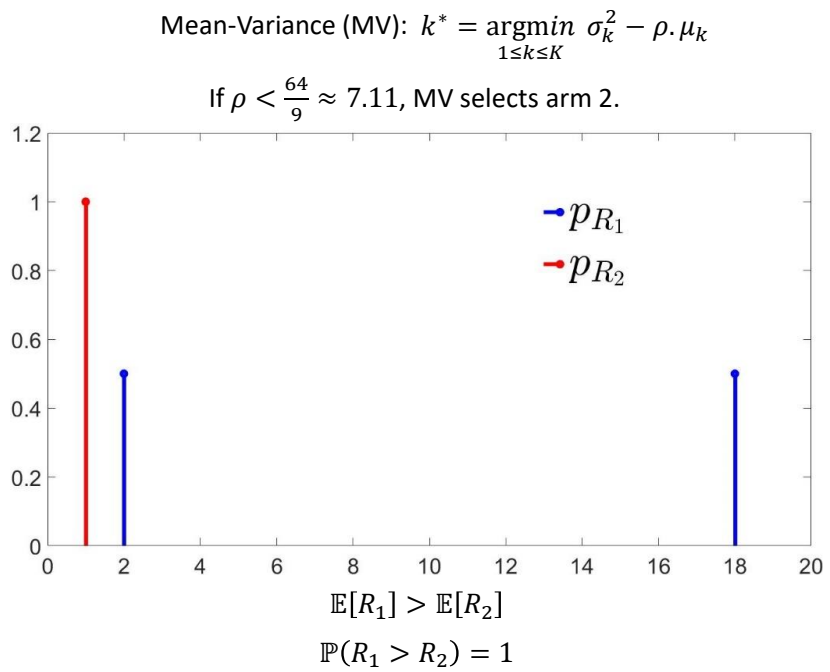

Figure 2.1: The example shows that mean-variance framework does not necessarily behave in a risk-averse manner.

the second arm if $\rho < 64/9$. On the other hand, the CVaR_α framework has a local view on the bottom of the support of the marginal distributions, so it misses the opportunities on the top part of the support. This fact is shown as an example in Figure 2.2, where the blue arm, R_1, rewards more than the red arm, R_2, with probability 0.81, but the CVaR_α selects the second arm if $\alpha < 0.2081$. In order to address these issues, we alternatively propose the following best arm identification algorithm for One-Time (Finite-time) Exploitation in a Multi-Armed Bandit problem (OTE/FTE-MAB algorithm) that has concrete mathematical support for its action and is hyper-parameter-free.

2.2 Risk-Averse Explore-Then-Commit Bandits with One/Finite-Time Exploitations

In this section, we propose the OTE-MAB and FTE-MAB algorithms. The OTE-MAB algorithm is a specific case of the FTE-MAB algorithm. Since the proof of theorem related to the FTE-MAB algorithm is notationally heavy,

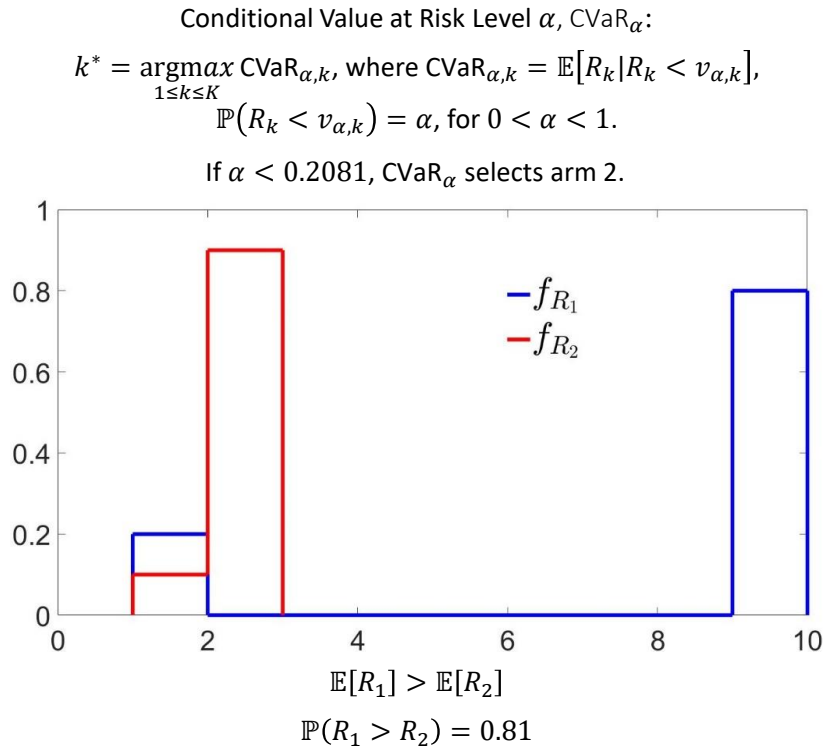

Figure 2.2: The example shows that the local view on the bottom of support of reward distributions in the CVaR$_\alpha$ framework misses the opportunities on the top part of the support.

the OTE-MAB algorithm is proposed first in Subsection 2.2.1 and the FTE-MAB algorithm is postponed to Subsection 2.2.2.

2.2.1 The OTE-MAB Algorithm

The OTE-MAB algorithm seeks to identify the arm that is most probable to reward the most for the case $M = 1$ as

$$k^* = \arg\max_k \mathbb{P}(R_k \geq \boldsymbol{R}_{-k}), \tag{2.3}$$

which is a specific case of Equation (2.1). For ease of notation, the M-notation is eliminated in this subsection.

Remark 1. *If there is any hard constraint on the minimum required reward in the one-time exploitation, c, the hard constraint can be concatenated to vector $\boldsymbol{R}_{-k}$ as $\boldsymbol{R}_{-k} = \{R_1, R_2, \ldots, R_{k-1}, R_{k+1}, \ldots, R_K, c\}$.*

Algorithm 1 The OTE-MAB Algorithm

Input $0 < \epsilon_r, \Delta p < 1$
choose $N \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$
Experimentation Phase:
  for $n = 1$ to N **do**
    $r_{k,n}$ is observed for all $k \in \mathcal{K}$
  end for
Calculate $\hat{p}_k = \frac{\sum_{n=1}^{N} \mathbb{1}\{r_{k,n} \geq \boldsymbol{r}_{-k,n}\}}{N}$
One-Time Exploitation:
  Play arm $\hat{k} = \arg\max_k \hat{p}_k$.

Since the joint reward distribution of the K arms are not known, the exact values of $p_k = \mathbb{P}(R_k \geq \boldsymbol{R}_{-k})$ are unknown. Hence, estimates of these probabilities, $\hat{p}_k$, are needed to be evaluated based on the observations in the experimentation phase as follows:

$$\hat{p}_k = \frac{\sum_{n=1}^{N} \mathbb{1}\{r_{k,n} \geq \boldsymbol{r}_{-k,n}\}}{N}, \tag{2.4}$$

where $\boldsymbol{r}_{-k,n} = \left(r_{1,n}, r_{2,n}, \ldots, r_{k-1,n}, r_{k+1,n}, \ldots, r_{K,n}\right)$. The OTE-MAB algorithm selects arm $\hat{k} = \arg\max_k \hat{p}_k$ as the best arm in terms of rewarding the most with the highest probability in one-time exploitation. The one-time exploitation regret for selecting arm $\hat{k}$, $r(\Delta p)$, which is a specific case of Definition 1, is

$$r(\Delta p) = \mathbb{P}\left(p_{k^*} - p_{\hat{k}} \geq \Delta p\right). \tag{2.5}$$

The OTE-MAB algorithm is summarized in Algorithm 1. The reason uniform exploration is utilized rather than a dynamic exploration in the pure exploration phase of Algorithm 1 is the following. The score of any arm is derived from the joint distribution of all arm rewards; as a result, stopping the exploration of any arm results in ceasing the concentration of the other arm scores. We next present a theorem on an upper bound of the minimum number of experiments needed to guarantee an upper bound on regret of Algorithm 1.

Theorem 1. *For any $0 < \epsilon_r, \Delta p < 1$, if all of the K arms are experimented jointly for $N \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$ times in the experimentation phase, the one-time exploitation regret is bounded by ϵ_r, i.e. $r(\Delta p) \leq \epsilon_r$.*

Refer to Appendix A.1 for the proof of Theorem 1.

According to Theorem 1 and using the law of total probability, the selected arm by Algorithm 1, $\hat{k}$, satisfies $\mathbb{E}\left[p_{\hat{k}}\right] \geq (1-\epsilon_r)\cdot(p_{k^*}-\Delta p)$ for any $0<\epsilon_r,\Delta p<1$, if all of the K arms are explored jointly in the experimentation phase for $N \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$ times. Furthermore, $p_{\hat{k}} \leq p_{k^*}$, so $p_{\hat{k}}$ can get arbitrarily close to p_{k^*} by increasing the number of pure explorations in the experimentation phase.

Let $p_{(1)}, p_{(2)}, \ldots, p_{(K)}$ be the ordered list of $p_1, p_2, \ldots, p_K$ in descending order. Note that arm (1) is actually arm k^* defined in Equation (2.3). Define the difference between the two maximum p_k's as $\Delta p^* = p_{(1)} - p_{(2)}$, where without loss of generality is assumed to be nonzero. Having the knowledge of Δp^* or a lower bound on it, a stronger notion of regret can be defined as

$$r = \inf_{\Delta p>0} r(\Delta p) = \mathbb{P}\left(\hat{k} \neq k^*\right), \tag{2.6}$$

and have the following corollary.

Corollary 1. *From the theoretical point of view, upon the knowledge of Δp^* or a lower bound on it, for any $0<\epsilon_r<1$, the regret defined in Equation (2.6) is bounded by ϵ_r, i.e. $r<\epsilon_r$, if all of the K arms are explored jointly for $N \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^{*2}}$ times.*

Remark 2. *If the K arms are independent, instead of estimating p_k by Equation (2.4), the following can be used:*

$$\hat{p}_k = \frac{\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\cdots\sum_{n_K=1}^{N}\mathbb{1}\{r_{k,n_k} \geq \boldsymbol{r}_{-k,n_{-k}}\}}{N^K}, \tag{2.7}$$

where $\boldsymbol{r}_{-k,n_{-k}} = \left(r_{1,n_1}, r_{2,n_2}, \ldots, r_{k-1,n_{k-1}}, r_{k+1,n_{k+1}}, \ldots, r_{K,n_K}\right)$. The above estimation can outperform the one in Equation (2.7), which is a promising future work. In the following, the challenge for obtaining a tighter confidence interval for estimates of p_k from Equation (2.7) versus Equation (2.4) is presented. For the case of dependent arms, there is an N-tuple containing the instantaneous observation of the K arm rewards as $(r_{1,n}, r_{2,n}, \ldots, r_{K,n})$ for $n \in \{1,2,\ldots,N\}$, which is used for estimation of $\hat{p}_k$ in Equation (2.4). On the other hand, for the case of independent arms, any of the N^K orderings of the N observations of the K arm rewards can be used for estimation of $\hat{p}_k$ as is done in Equation (2.7). However, $\left(\hat{p}_k - \frac{a}{2\sqrt{N^K}}, \hat{p}_k + \frac{a}{2\sqrt{N^K}}\right)$ cannot

be used as confidence interval with confidence level $1-2e^{-\frac{a^2}{2}}$. *The reason is that, although* $\hat{p}_k$ *is derived from* N^K *samples, not all those samples are independent, but exactly* N *of the* N^K *samples are independent. In fact, the observed independent rewards can be classified as* N*-tuples of the* K *arm rewards with independent elements in* $N^{k-1}\times(N-1)^{k-1}\times\cdots\times 1^{k-1}=(N!)^{K-1}$ *different ways. None of such* N*-tuples has any priority over the other ones to estimate* p_k*, so* $\hat{p}_k$ *can be computed based on any of the* N*-tuples. The estimate of* p_k *derived from any of those* N*-tuples is in* $\left(p_k-\frac{a}{2\sqrt{N}},p_k+\frac{a}{2\sqrt{N}}\right)$ *with probability at least* $1-2e^{-\frac{a^2}{2}}$*, so the average of those estimations is again in the mentioned interval with probability at least* $1-2e^{-\frac{a^2}{2}}$. *Note that the average of estimates of* p_k *derived from all of the* $(N!)^{K-1}$ *different* N*-tuples is equal to* $\hat{p}_k$ *derived from Equation* (2.7) *due to the following reason. An element of an* N*-tuple is repeated for* $((N-1)!)^{K-1}$ *times in all* N*-tuples. Hence, averaging over the* $\frac{(N!)^{K-1}\cdot N}{((N-1)!)^{K-1}}=N^K$ *number of distinct elements of* N*-tuples results in the same answer as the case of averaging the estimates of* p_k *derived from all of* $(N!)^{K-1}$ *different* N*-tuples. As a result,* $\frac{a}{2\sqrt{N}}$ *can be used as the half width of the confidence interval for estimators obtained from Equation* (2.7) *for independent arms.*

2.2.2 The FTE-MAB Algorithm

Consider the case where an arm is going to be exploited for finite number of times, $M<\infty$. The best arm for M-time exploitations is defined in Equation (2.1). Since the joint reward distribution is unknown, p_k^M's are needed to be estimated based on observations in the pure exploration phase. Define the vector $\mathcal{R}_k^M$, that is not unique, with cardinality $\lfloor\frac{N}{M}\rfloor$ as

$$\begin{aligned}\mathcal{R}_k^M=\Bigg\{&\sum_{n\in S_i} r_{k,n} \text{ for } 1\leq i\leq \lfloor\frac{N}{M}\rfloor \text{ s.t. } S_i,S_j\subseteq\{1,\ldots,N\},\\ &|S_i|=|S_j|=M, \text{ and } S_i\cap S_j=\varnothing, 1\leq\forall i\neq j\leq\lfloor\frac{N}{M}\rfloor\Bigg\},\end{aligned} \tag{2.8}$$

where $r_{k,j}^M$ for $1\leq j\leq\lfloor\frac{N}{M}\rfloor$ are the different elements of $\mathcal{R}_k^M$. Let the set S_i corresponding to $r_{k,j}^M$ be used for generating $r_{k',j}^M$ for all $k'\in\mathcal{K}$. Let $\hat{p}_k^M$ be

Algorithm 2 The FTE-MAB Algorithm

Input $0 < \epsilon_r, \Delta p < 1$ and $M \geq 1$
choose N such that $\lfloor \frac{N}{M} \rfloor \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$
Experimentation Phase:
  for $n = 1$ to N **do**
    $r_{k,n}$ is observed for all $k \in \mathcal{K}$
  end for
Let $\mathcal{R}_k^M = \Big\{ \sum_{n \in S_i} r_{k,n}$ for $1 \leq i \leq \lfloor \frac{N}{M} \rfloor$ s.t. $S_i, S_j \subseteq \{1, 2, \ldots, N\}, |S_i| = |S_j| = M,$ and $S_i \cap S_j = \varnothing, 1 \leq \forall i \neq j \leq \lfloor \frac{N}{M} \rfloor \Big\}$, where $r_{k,j}^M$ for $1 \leq j \leq \lfloor \frac{N}{M} \rfloor$ are the different elements of $\mathcal{R}_k^M$. Let the set S_i corresponding to $r_{k,j}^M$ be used for generating $r_{k',j}^M$ for all $k' \in \mathcal{K}$.
Calculate $\hat{p}_k^M = \frac{\sum_{j=1}^{\lfloor \frac{N}{M} \rfloor} \mathbb{1}\{r_{k,j}^M \geq \boldsymbol{r}_{-k,j}^M\}}{\lfloor \frac{N}{M} \rfloor}$
M-Time Exploitation:
  Play arm $\hat{k} = \arg\max_k \hat{p}_k^M$ for M times.

the estimate of p_k^M that can be computed as

$$\hat{p}_k^M = \frac{\sum_{j=1}^{\lfloor \frac{N}{M} \rfloor} \mathbb{1}\{r_{k,j}^M \geq \boldsymbol{r}_{-k,j}^M\}}{\lfloor \frac{N}{M} \rfloor}. \tag{2.9}$$

The FTE-MAB algorithm selects arm $\hat{k} = \arg\max_k \hat{p}_k^M$ for M-time exploitations. This algorithm is summarized in Algorithm 2. We next present a theorem for an upper bound of the minimum number of experiments needed to guarantee an upper bound on regret of Algorithm 2 which is the generalization of Theorem 1.

Theorem 2. *For any $0 < \epsilon_r, \Delta p < 1$, if all of the K arms are explored jointly for N times in the experimentation phase such that $\lfloor \frac{N}{M} \rfloor \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$, the finite-time exploitation regret is bounded by ϵ_r, i.e. $r_M(\Delta p) \leq \epsilon_r$.*

The proof of Theorem 2 is similar to that of Theorem 1, which can be found in Appendix A.2.

Let $p_{(1)}^M, p_{(2)}^M, \ldots, p_{(K)}^M$ be the ordered list of p_1^M, p_2^M, $\ldots$, p_K^M in descending order. Note that arm (1) is actually arm k^* defined in Equation (1). Define the difference between the two maximum p_k^M's as $\Delta p_M^* = p_{(1)}^M - p_{(2)}^M$, where without loss of generality is assumed to be nonzero. Having the knowledge

of Δp_M^* or a lower bound on it, a stronger notion of regret can be defined as

$$r_M = \inf_{\Delta p > 0} r_M(\Delta p) = \mathbb{P}\left(\hat{k} \neq k^*\right), \tag{2.10}$$

and the following corollary follows.

Corollary 2. *From the theoretical point of view, upon the knowledge of Δp_M^* or a lower bound on it, for any $0 < \epsilon_r < 1$, the regret defined in Equation (2.10) is bounded by ϵ_r, i.e. $r_M < \epsilon_r$, if all of the K arms are explored jointly for N times, where $\lfloor \frac{N}{M} \rfloor \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{{\Delta p_M^*}^2}$.*

Remark 3. *We note that the need for the number of samples to scale linearly with M in Theorem 2 may seem sub-optimal at first. This is a consequence of having a distribution-independent statement of the theorem. We provide an example in Section 2.4 that shows the linear scaling for $M = 2$. If M converges to infinity, the problem becomes the classical multi-armed bandit problem since $\arg\max_k \mathbb{P}(R_k^M \geq \boldsymbol{R}_{-k}^M)$ is the same as $\arg\max_k \mathbb{P}\left(\frac{R_k^M}{M} \geq \frac{\boldsymbol{R}_{-k}^M}{M}\right)$ and due to the law of large numbers $\frac{R_k^M}{M} \to \mu_k$ as $M \to \infty$. Hence, the FTE-MAB algorithm selects the arm with maximum expected reward if the arm is going to be exploited for infinitely many times and the cumulative reward is desired to be maximized.*

Remark 4. *Let $\mathcal{R}_k^M = \left\{ \sum_{n \in S_\mathcal{K}} r_{k,n} \text{ s.t. } S_\mathcal{K} \subseteq \{1, 2, \ldots, N\} \text{ and } |S_\mathcal{K}| = M \right\}$, where $r_{k,j}^M$ for $1 \leq j \leq \binom{N}{M}$ are the different elements of $\mathcal{R}_k^M$. Let the set $S_\mathcal{K}$ corresponding to $r_{k,j}^M$ be used for generating $r_{k',j}^M$ for all $k' \in \mathcal{K}$. The estimates of p_k^M can be calculated as*

$$\hat{p}_k^M = \frac{\sum_{j=1}^{\binom{N}{M}} \mathbb{1}\{r_{k,j}^M \geq \boldsymbol{r}_{-k,j}^M\}}{\binom{N}{M}} \tag{2.11}$$

or if the K arms are independent, p_k^M can be estimated as

$$\hat{p}_k^M = \frac{\sum_{j_1=1}^{\binom{N}{M}} \sum_{j_2=1}^{\binom{N}{M}} \cdots \sum_{j_K=1}^{\binom{N}{M}} \mathbb{1}\{r_{k,j_k}^M \geq \boldsymbol{r}_{-k,j_{-k}}^M\}}{\binom{N}{M}^K}. \tag{2.12}$$

An interesting future work is to obtain a tighter confidence interval for estimates of p_k^M from Equation (2.11) or Equation (2.12) versus Equation (2.9).

2.3 A Cost-Based Analysis for Risk-Averse Explore-Then-Commit Two-Armed Bandits

Up to this point, the experimentation cost is not considered. However, if experimentation is time-consuming, there is a cost in postponing the exploitation of the best identified arm. For example, for more experimentation, a patient receives medication by delay or an investor keeps his/her money on hold with zero interest, both of which incur costs. As explained in Section 2.1, let such a cost be formulated by a function $C(.)$, where $C(n)$ is the incurred cost of n joint experiments of all arms. Then, a trade-off between cost and regret emerges, where increasing the number of explorations decreases regret, but increases the incurred cost. Such a trade-off can be formalized using a hyper-parameter by solving

$$N^* = \arg\min_n \; C(n) + \alpha \cdot r^*(n, p_{k^*}), \tag{2.13}$$

where α characterizes the cost-regret trade-off, $r^*(n, p_{k^*}) = \mathbb{P}(\hat{k} \neq k^*)$, defined in Equation (2.5) when $\Delta p = 0$, is the regret when n experiments are done, and $p_{k^*} = \max\left(\mathbb{P}(R_1 \geq R_2), \mathbb{P}(R_2 \geq R_1)\right)$, which is unknown. Define $Cr(n, p) = C(n) + \alpha \cdot r^*(n, p)$ as the cost-regret function. Note that upon the knowledge of p_{k^*}, the regret can be formulated as

$$\begin{aligned} r^*(n, p_{k^*}) = &\sum_{i=\lfloor \frac{n}{2} \rfloor + 1}^{n} \binom{n}{i} \cdot (1 - p_{k^*})^i \cdot p_{k^*}^{n-i} \\ &+ \frac{1}{2} \cdot \binom{n}{\frac{n}{2}} \cdot (1 - p_{k^*})^{\frac{n}{2}} \cdot p_{k^*}^{\frac{n}{2}} \cdot \mathbb{1}\{n \text{ is even}\}. \end{aligned} \tag{2.14}$$

Deriving regret from the above equation meets simulation-based results for regret of the OTE-MAB algorithm that is plotted in Figure 2.5 which is presented in Section 2.4. Figure 2.3 shows the cost-regret function $Cr(n, p_{k^*}) = C(n) + \alpha \cdot r^*(n, p_{k^*})$ under Example 1 for $C(n) = \frac{n}{10000}, \alpha = 1$, and when the parameter p_{k^*} is known. As shown, the cost-regret function is minimized at $N^* = 168$. Note that the parameter p_{k^*} is unknown, which raises questions on how confident one can be on finding an estimate of N^* based on an estimate of p_{k^*} which is discussed in more detail below.

After n_e number of joint explorations of arms, denote the estimates of

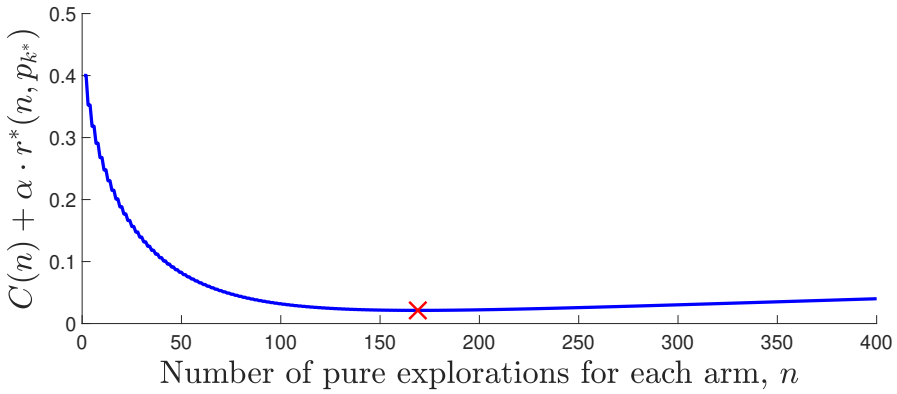

Figure 2.3: Cost-regret trade-off is addressed by minimizing a linear combination of cost and regret.

$p_1 = \mathbb{P}(R_1 \geq R_2)$ and $p_2 = \mathbb{P}(R_2 \geq R_1)$ by $\hat{p}_1(n_e)$ and $\hat{p}_2(n_e)$ that are derived from Equation (2.4). The parameter n_e should not be confused with sample iteration that is denoted by $n \in \{1, 2, 3, 4, \dots\}$. After the n_e observations of the joint arm rewards, the regret function $r^*(n, p_{k^*})$ can be estimated as $r^*(n, \hat{p}^*(n_e))$, where $\hat{p}^*(n_e) = \max(\hat{p}_1(n_e), \hat{p}_2(n_e))$, and the optimal number of experiments, N^*, can be estimated by a confidence level as

$$\hat{N}^*(n_e) = \arg\min_n \; C(n) + \alpha \cdot r^*(n, \hat{p}^*(n_e)). \tag{2.15}$$

As a complementary method, we suggest to use the confidence interval of $\hat{p}^*(n_e)$ in order to present an interval, $\mathcal{I}(n_e)$, that includes the optimal stopping point, N^*, with a confidence level. It is proved later that the interval $\mathcal{I}(n_e)$ shrinks towards N^* as n_e increases. For a confidence level $1 - 2e^{-\frac{a^2}{2}}$, the estimate of p_{k^*}, $\hat{p}^*(n_e)$, has the property that

$$\begin{aligned} &P\left(p_{k^*} \in \left(\max\left\{\hat{p}^*(n_e) - \frac{a}{2\sqrt{n_e}}, 0.5\right\}, \min\left\{\hat{p}^*(n_e) + \frac{a}{2\sqrt{n_e}}, 1\right\}\right)\right) \\ \geq &1 - 2e^{-\frac{a^2}{2}}. \end{aligned} \tag{2.16}$$

Denote the lower and upper bounds of the confidence interval as $\hat{p}_l^*(n_e) = \max\left\{\hat{p}^*(n_e) - \frac{a}{2\sqrt{n_e}}, 0.5\right\}$ and $\hat{p}_u^*(n_e) = \min\left\{\hat{p}^*(n_e) + \frac{a}{2\sqrt{n_e}}, 1\right\}$, respectively. Let $Cr_l(n, n_e) \triangleq Cr(n, \hat{p}_u^*(n_e))$ and $Cr_u(n, n_e) \triangleq Cr(n, \hat{p}_l^*(n_e))$. It is shown later that $Cr(n, p)$ is decreasing with respect to $0.5 \leq p \leq 1$, and that is why $Cr_l(n, n_e)$ is associated with $\hat{p}_u^*(n_e)$ and $Cr_u(n, n_e)$ with $\hat{p}_l^*(n_e)$ so that $Cr_l(n, n_e) \leq Cr_u(n, n_e)$ for any $n \in \{1, 2, 3, \dots\}$. Let the minimizer of

Algorithm 3 The c-OTE-MAB Algorithm

Input $a > \sqrt{2\ln(2)}$ and $n_e \geq 1$ experiments for the joint arm rewards as r_{k,n_k} for $k \in \{1,2\}$ and $1 \leq n_k \leq n_e$.

Parameter and function estimations:

Calculate $\hat{p}_1(n_e) = \frac{\sum_{n=1}^{n_e} \mathbb{1}\{r_{1,n} \geq \boldsymbol{r}_{2,n}\}}{N}$

$\hat{p}^*(n_e) = \max\{\hat{p}_1(n_e), 1 - \hat{p}_1(n_e)\}$

$\hat{p}_l^*(n_e) = \max\left\{\hat{p}^*(n_e) - \frac{a}{2\sqrt{n_e}}, 0.5\right\}$

$\hat{p}_u^*(n_e) = \min\left\{\hat{p}^*(n_e) + \frac{a}{2\sqrt{n_e}}, 1\right\}$

$Cr_l(n, n_e) \triangleq Cr(n, \hat{p}_u^*(n_e))$

$Cr_u(n, n_e) \triangleq Cr(n, \hat{p}_l^*(n_e))$

$N_u^* = \arg\min_n Cr_u(n, n_e)$

Exploration stopping iteration $\hat{N}^*(n_e)$ **and stopping interval** $\mathcal{I}(n_e)$**:**

$\hat{N}^*(n_e) = \arg\min_n C(n) + \alpha \cdot r^*(n, \hat{p}^*(n_e))$

$\mathcal{I}(n_e) = \{n : Cr_l(n, n_e) \leq Cr_u(N_u^*, n_e)\}$

the upper-bound function be $N_u^* = \arg\min_n Cr_u(n, n_e)$, then the following interval is proposed that includes the optimal stopping point and shrinks towards it as n_e increases with the aforementioned confidence level:

$$\mathcal{I}(n_e) = \{n : Cr_l(n, n_e) \leq Cr_u(N_u^*, n_e)\}. \tag{2.17}$$

Note that $\arg\min\{Cr_u(n, n_e)\}$ can have multiple solutions, so N_u^* is not necessarily a unique number. Hence, throughout this chapter, we set a convention as $Cr_u(N_u^*, n_e) = Cr_u(n, n_e)$ for any $n \in N_u^*$. The cost-based algorithm, called the c-OTE-MAB algorithm, discussed in this section is summarized in Algorithm 3. The pictorial expression of this algorithm is depicted in Figure 2.4 for a non-monotonic cost function. In this figure, the minimum of each plot is shown by a cross sign. In the following, we present a theorem on $\hat{N}^*(n_e)$ and $\mathcal{I}(n_e)$ given by Algorithm 3.

Theorem 3. *Possessing n_e number of joint experiments for the two arms and assuming that $p_{k^*} \in [0.5 + \epsilon_p, 1]$ when $\epsilon_p \in (0, 0.5]$ is an unknown parameter, we have*

$$\begin{aligned} &Cr\left(\hat{N}^*(n_e), p_{k^*}\right) - Cr\left(N^*, p_{k^*}\right) \\ \leq &\frac{D_p}{2\sqrt{n_e}} + \Delta Cr(\hat{N}^*(n_e), n_e) \xrightarrow{n_e \to \infty} 0 \end{aligned} \tag{2.18}$$

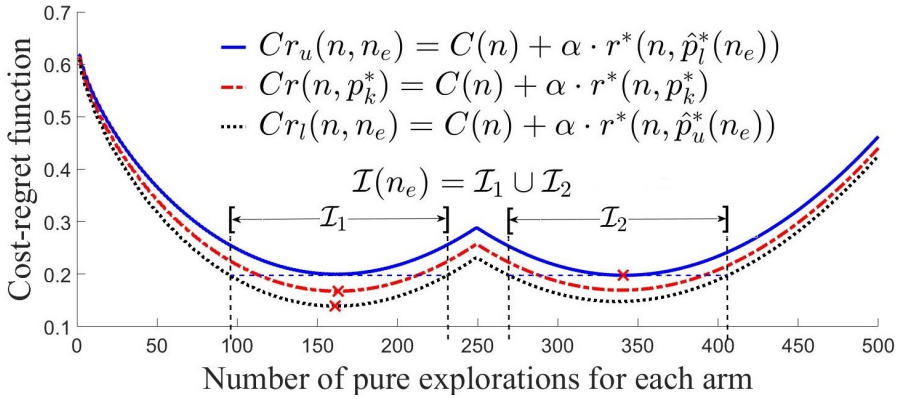

Figure 2.4: The pictorial expression of the stopping interval $\mathcal{I}(n_e)$ in Algorithm 3.

and

$$\max_{n \in \mathcal{I}(n_e)} \left(Cr(n, p_{k^*}) - Cr(N^*, p_{k^*}) \right) \leq \frac{D_p}{\sqrt{n_e}} \tag{2.19}$$

with confidence level $1 - 2e^{-\frac{a^2}{2}}$*, where* $\hat{N}^*(n_e)$*,* N^**, and* $\mathcal{I}(n_e)$ *are defined in Equations* (2.15)*,* (2.13)*, and* (2.17)*, respectively,* D_p *is a constant as* $D_p = \frac{a \cdot \alpha \cdot 2^{\left(4\delta_p + 1 - \frac{1}{2 \ln 2}\right)}}{\sqrt{2\delta_p \ln 2}}$*, where* $\delta_p = \frac{1}{2}\left(-2 - \log_2(0.5 + \epsilon_p) - \log_2(0.5 - \epsilon_p)\right) > 0$*, and* $\Delta Cr(n, n_e) = \frac{a \cdot \alpha \cdot \sqrt{n+2} \cdot 2^{-\delta_p \cdot (n-2)}}{\sqrt{n_e}} \leq \frac{D_p}{2\sqrt{n_e}}$ *for any* $n \in \{1, 2, 3, \dots\}$.

Refer to Appendix A.3 for the proof of Theorem 3.

Using the proof results of Theorem 3, the following corollaries are followed.

Corollary 3.

$$\lim_{n_e \to \infty} \mathbb{E}\left[\hat{N}^*(n_e)\right] = N^*. \tag{2.20}$$

Refer to Appendix A.4 for the proof of Corollary 3. We note that in practice $\mathbb{E}\left[\hat{N}^*(n_e)\right]$ converges to N^* relatively fast when the exploration cost is relatively small as is shown by simulation in Section 2.4.

Corollary 4. *The set of optimal stopping points* N^* *defined in Equation* (2.13) *is a subset of the set* $\mathcal{I}(n_e)$ *defined in Equation* (2.17) *with the associated confidence level, i.e.* $N^* \subseteq \mathcal{I}(n_e)$ *with confidence level* $1 - 2e^{-\frac{a^2}{2}}$*. Furthermore,* $\mathcal{I}(n_e) = N^*$ *with the mentioned confidence level for* $n_e > \frac{D_p^2}{\left(Cr(N^*, p_{k^*}) - \min\limits_{n \notin N^*} Cr(n, p_{k^*})\right)^2}$.

For the proof of Corollary 4 refer to Appendix A.5.

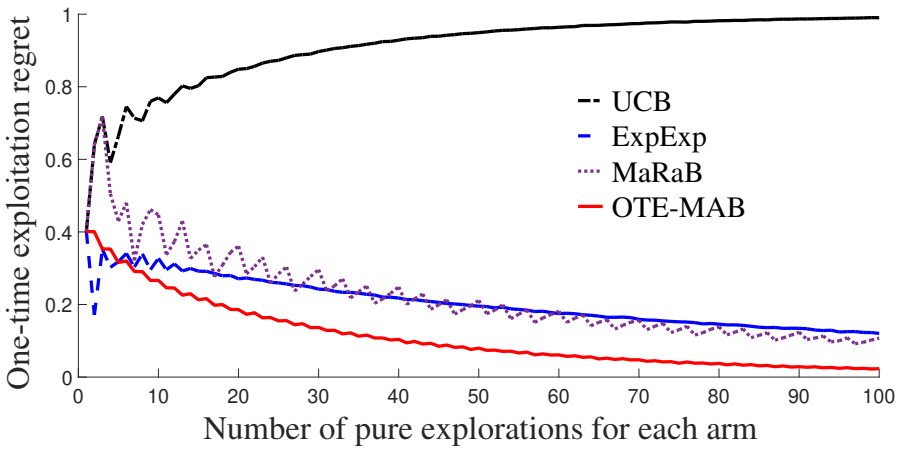

Figure 2.5: Comparison of regret for OTE-MAB against the state-of-the-art algorithms for Example 1.

2.4 Simulation Results

In this section, numerical simulations validating the theoretical results presented in this chapter are reported. The proposed OTE-MAB algorithm is compared with the Upper Confidence Bound (UCB) [184], Mean-Variance based ExpExp [17], and CVaR_α based MaRaB [22] algorithms. Consider two arms with the reward distributions given in Example 1. The regret defined in Equation (2.6) versus the number of pure explorations for each arm, N, is averaged over 100,000 runs. The result is plotted in Figure 2.5 and as it is shown, OTE-MAB outperforms the state-of-the-art algorithms for the purpose of risk-aversion in terms of the regret defined in this chapter. Note that the UCB algorithm aims at selecting an arm that maximizes the expected received reward, but in Example 1, the arm with higher expected reward is less probable to have the highest reward for one-time exploitation, which is why the UCB algorithm performs poorly in this example. However, in the following example where the arm that rewards more on expectation is also more probable to reward more, the UCB, ExpExp, and MaRaB algorithms perform as well as the OTE-MAB algorithm.

Example 2. *Consider two arms with the following unknown independent reward distributions:*

$$f_1(u) = \alpha e^{-0.5(u-2)^2} \cdot \mathbb{1}\{0 \leq u \leq 10\}$$
$$f_2(u) = \beta e^{-0.5(u-1)^2} \cdot \mathbb{1}\{0 \leq u \leq 10\},$$

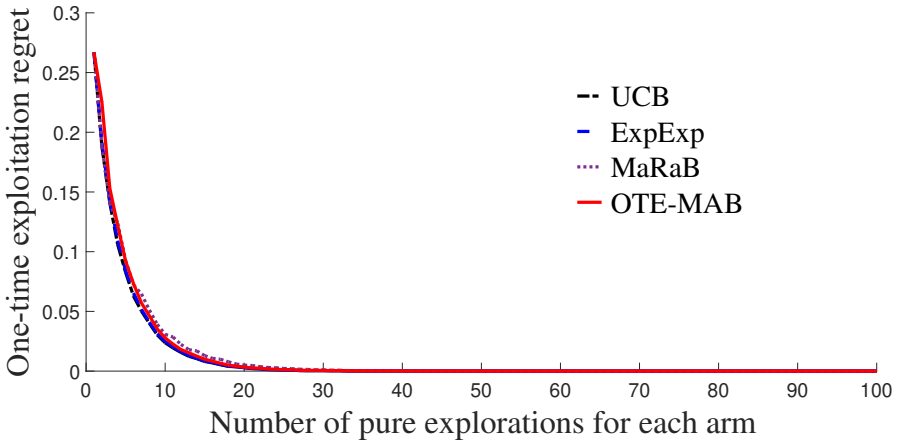

Figure 2.6: Comparison of regret for OTE-MAB against the state-of-the-art algorithms for Example 2.

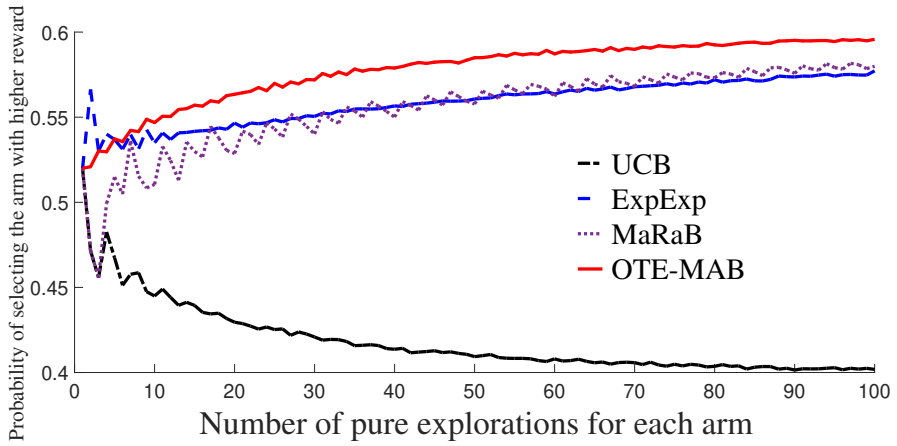

Figure 2.7: Comparison of probability of selecting the arm with higher reward for OTE-MAB against the state-of-the-art algorithms for Example 1.

where α and β are constants so that the two probability distribution functions integrate to one.

Note that in example 2, $\mathbb{E}[R_1] > \mathbb{E}[R_2]$ and $\mathbb{P}(R_1 \geq R_2) > 0.5$. For this scenario, the regret defined in Equation (2.6) versus the number of pure explorations for each arm, N, averaged over 100,000 runs is plotted in Figure 2.6.

In another experiment, the multi-armed bandit is simulated for Example 1, where the probability that the selected arm has the higher reward is calculated over 500,000 runs for different algorithms. The result is shown in Figure 2.7. This result confirms the motivation of our study on risk-averse

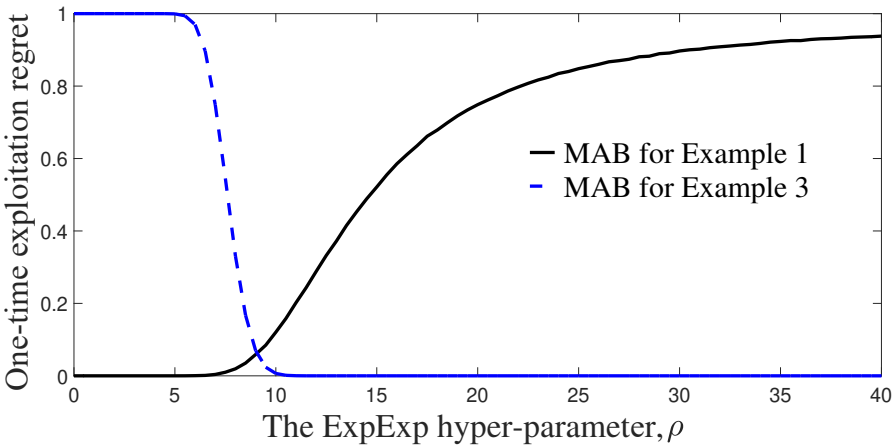

Figure 2.8: Regret of the ExpExp algorithm versus the hyper-parameter ρ for two examples.

finite-time exploitations in multi-armed bandits.

In the above comparison of OTE-MAB with state-of-the-art algorithms, three different choices of hyper-parameters for the ExpExp and MaRaB algorithms are tested and the best performance is presented. However, note that the performances of these algorithms depend on the choice of hyper-parameter. In Figure 2.8, the sensitivity of the performance of the ExpExp algorithm with respect to the choice of hyper-parameter ρ is depicted for Example 1 and a third example where the variance of the best arm is larger than the variance of the arm with lower expected reward. The two plots are the averaged regret over 100,000 runs versus the value of ρ for the ExpExp algorithm for two different multi-armed bandit problems when $N = 100$. As depicted in Figure 2.8, a choice of ρ can be good for one multi-armed bandit problem, but not good for another one. Due to our observations, the sensitivity of the MaRaB algorithm to its hyper-parameter can even be more complex. Figure 2.9 depicts the averaged regret over 100,000 runs versus the value of MaRaB hyper-parameter, α, when $N = 100$. This figure is plotted for Example 1 and a fourth example where reward of the first arm has a truncated normal distribution with mean three and variance two over the interval $[0, 10]$ and the second arm is the same as the one in Example 1.

In another experiment, the minimum number of explorations needed to guarantee a bound on regret is compared for two cases of one-time and two-time exploitations. Theorems 1 and 2 suggest that for given K, ϵ_r, and $\Delta p^* = \Delta p^*_M$, the upper bound of minimum number of explorations needed for M-

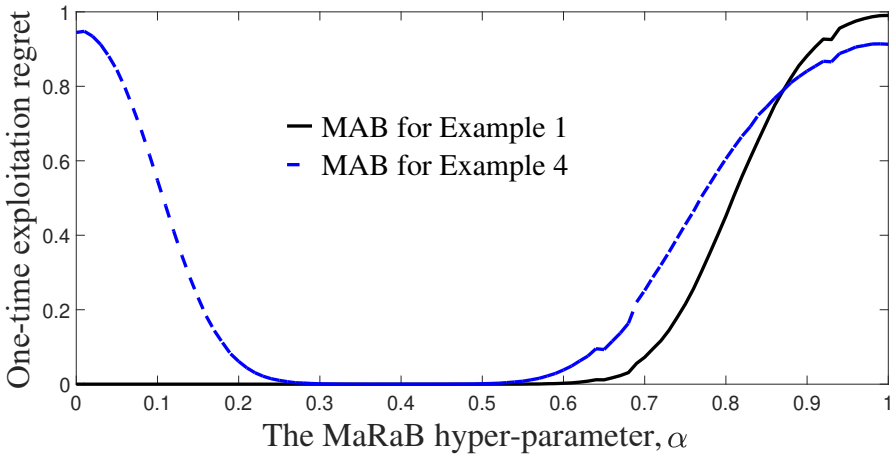

Figure 2.9: Regret of the MaRaB algorithm versus the hyper-parameter α for two examples.

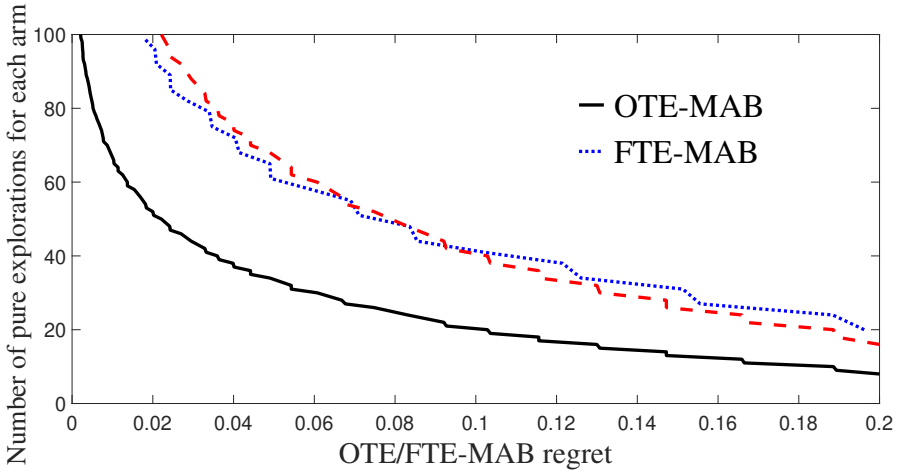

Figure 2.10: The minimum number of explorations needed to guarantee a bound on regret for two cases of one-time and two-time exploitations.

time exploitations to guarantee that the regret is bounded by ϵ_r is M times that of one-time exploitation. We design two examples of two-armed bandits such that $\Delta p^* = \Delta p_2^* = 0.28$ and plot the minimum number of explorations to guarantee bounded regret by ϵ_r in Figure 2.10. The dashed line is the plot of the OTE-MAB algorithm multiplied by two which is close to the one related to the FTE-MAB algorithm for two-armed bandits. This observation provides support to our theoretical results.

Theorem 3 states that the expected of $\hat{N}^*(n_e)$ converges to N^* as the number of explorations goes to infinity. In practice, it is observed that $\mathbb{E}\left[\hat{N}^*(n_e)\right]$

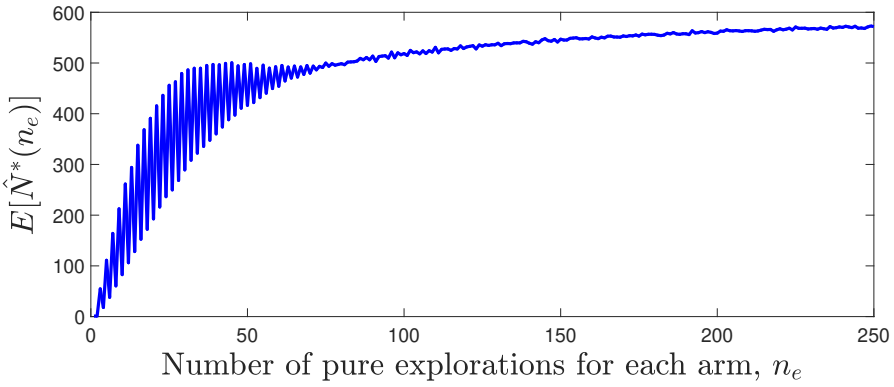

Figure 2.11: $\mathbb{E}\left[\hat{N}^*(n_e)\right]$ versus n_e for $p_{k^*} = 0.54$ and $\alpha = 1$, where $N^* = 587$.

converges relatively fast. The distribution of $\hat{p}^*(n_e)$ for an odd n_e is

$$\begin{aligned} &\mathbb{P}\left(\hat{p}^*(n_e) = 1 - \frac{j}{n_e}\right) \\ =&\binom{n_e}{n_e - j} \cdot p_{k^*}^{n_e - j} \cdot (1 - p_{k^*})^j + \binom{n_e}{j} \cdot p_{k^*}^j \cdot (1 - p_{k^*})^{n_e - j}, \end{aligned} \tag{2.21}$$

for $0 \leq j \leq \frac{n_e - 1}{2}$, and the distribution of $\hat{p}^*(n_e)$ for an even n_e is

$$\begin{aligned} &\mathbb{P}\left(\hat{p}^*(n_e) = 1 - \frac{j}{n_e}\right) \\ =&\binom{n_e}{n_e - j} \cdot p_{k^*}^{n_e - j} \cdot (1 - p_{k^*})^j + \binom{n_e}{j} \cdot p_{k^*}^j \cdot (1 - p_{k^*})^{n_e - j}, \end{aligned} \tag{2.22}$$

for $0 \leq j < \frac{n_e}{2} - 1$ and $\mathbb{P}\left(\hat{p}^*(n_e) = \frac{1}{2}\right) = \binom{n_e}{\frac{n_e}{2}} \cdot p_{k^*}^{\frac{n_e}{2}} \cdot (1 - p_{k^*})^{\frac{n_e}{2}}$. Equations (2.15), (2.21), and (2.22) are used to plot Figure 2.11 that shows $\mathbb{E}\left[\hat{N}^*(n_e)\right]$ versus n_e for $p_{k^*} = 0.54$ and $\alpha = 1$. Note that the optimal stopping point when $p_{k^*} = 0.54$ and $\alpha = 1$ is $N^* = 587$ and $\mathbb{E}\left[\hat{N}^*(n_e)\right]$ is approaching this value as depicted in the figure.

Chapter 3

RISK-AVERSE EQUILIBRIUM FOR STOCHASTIC GAMES

Since the seminal work of von Neumann and Morgenstern [28], the term *rational* has become synonymous with expected utility maximization. Whether in game theoretic situations or simply decision-making under uncertainty, the only agent who can be considered rational is the one who attempts to maximize their mean utility, no matter how many trials will likely be necessary for the realized value to resemble the expected value. However, consider an agent faced with multiple options, one of which is an opportunity with maximum expected utility, but it will bankrupt them with high probability if it fails. In the event of failure, consider that the lack of funds will severely limit any future options the agent may have. For such an agent the fact that the opportunity has maximum expected value among the options cannot be the only relevant factor in deciding whether to pursue the opportunity. If the opportunity does not lead to success, the agent will not be able to pursue any later actions, as they will not have the funds necessary to do so. As a result, players should not solely rely on factors such as expected utility and must instead also consider the probability of success for the opportunity.

This observation applies to almost all stochastic decision-making situations, including competitive situations best modeled through game theory. To see this, consider a market composed of only a few large firms and a smaller firm considering how to compete with large firms or whether to even enter the market. We take as our example the smartphone industry, in which large companies such as Apple, Samsung, Google, LG, Motorola, Amazon, and Microsoft have all competed in recent years. While Apple and Samsung are market leaders at the time of writing, both have undergone expensive setbacks. Apple's iPhone 5 was widely criticized due to issues with the Apple Maps application and Samsung had to recall its Galaxy Note 7 due to its batteries catching fire, costing an estimated 3 billion USD [185], in what may have been an attempt to improve on the criticized battery life of their Galaxy

S6. Similarly, Google's original Nexus line of phones dropped in popularity to the point where the company went to the expense of creating a new line of Pixel phones rather than continuing the Nexus. Amazon and Microsoft were forced out of the market entirely, with Amazon's Fire Phone lasting just over a year (July 2014 - August 2015) between release and the cessation of production, causing a loss of at least 170 million USD for Amazon's 2014 Q3 alone [186]. Microsoft meanwhile acquired Nokia for 7.2 billion USD in an attempt to become more competitive in the market [187], but ceased mobile device production entirely only a few years later.

Despite the cost of the setbacks mentioned above, each of these companies is still valuable with Apple and Microsoft having market caps of over 1 trillion USD at the time of writing and Amazon recently passing that milestone as well. Samsung is worth approximately 300 billion USD at the time of writing, and while they are smaller LG and Motorola are quite valuable as well, worth approximately 14.5 billion and 30 billion USD, respectively. Because of their size, each of these companies was able to take risks to compete with each other which, although expected to end in a positive outcome, resulted in expensive losses. Indeed, Microsoft currently appears to be preparing for another attempt to enter the smartphone market with the Surface Duo. In other words, these companies are still able to compete with each other by making products which maximize their expected values because they are large enough that they can afford to wait for the law of large numbers to take effect. This allows their competition to be modeled through a traditional game theoretic framework.

In contrast, consider a company with a smaller valuation, say 500 million USD, deciding whether to compete in the smartphone market. If such a company attempted to do so, it would have to commit most if not all of its resources to the attempt. Even if such a strategy has a large positive expected value, it has a large risk of bankrupting the company, as seen with the scale of the losses incurred by Samsung, Amazon, and Microsoft. More generally, firms in markets where the cost of competition is a significant portion of the value of the firm itself must consider more than just maximizing their expected value. A misstep in such a setting means that the firm is out of the market and unable to compete further. This highlights an important facet of competition with random or unknown variables; i.e., it is not just the expected value of a strategy that is important, it is how many times you

get to compete.

In this chapter, we build a new framework to apply this observation to game theoretic situations. We use a risk-averse best-response approach for incomplete information games drawn from known distributions in which players engage once or for a given finite number of times. Because of the finite number of times that players engage in these games, given the strategies of all other players, expected utility may not be a suitable metric for a player to attempt to maximize. Instead, we formulate a new definition of a risk-averse best response, where given the strategies of all other agents, an agent chooses to play the strategy that is most likely to have the highest utility in a single realization of the stochastic game. We show that the risk-averse equilibrium based on the mentioned probability statement can be found by realizing the Nash equilibrium of a new game whose payoffs are derived from the probability distributions of the payoffs of the original game. While the mathematical particulars of this definition will be discussed in Section 3.2, conceptually it can best be understood through the lens of prospect theory.

In its most basic form, prospect theory states that consumers prefer choices with lower volatility, even when this results in lower expected utility. An excellent example of this is retirement planning where there are many highly volatile assets which in expectation provide a large return on investment, but which also have a high chance of dropping in value due to their volatility. Most individuals try to avoid investing too much in these assets, receiving a lower average return in order to avoid the chance of a significant loss. Similarly, a risk-averse best response as we have loosely defined it so far would possibly limit the expected return of assets in order to maximize the probability of making the most profit.

The rest of this chapter is organized as follows. The problem statement of stochastic games is provided in Section 3.1. Section 3.2 provides the formal mathematical definition of the proposed risk-averse equilibrium, with several subsequent sections detailing topics such as equilibrium properties (Section 3.3), computation (Section 3.4), and worked-out examples (Section 3.5). Section 3.6 considers finite-time commit games and how the risk-averse equilibria shift as the number of times the games are played increases. Section 3.7 compares the classical Nash equilibrium and the proposed risk-averse equilibrium through simulation. The concluding remarks as well as future directions in which to advance this research are provided in Chapter 6.

3.1 Problem Statement of Stochastic Games

Consider a game that consists of a finite set of N players, $[N] := \{1, 2, \ldots, N\}$, where player $i \in [N]$ has a set of possible pure strategies (or actions, used interchangeably) denoted by S_i. A pure strategy profile, which is one pure strategy for each player in the game, is denoted by $\boldsymbol{s} = (s_1, s_2, \ldots, s_N)$, where $s_i \in S_i$ is the pure strategy of player $i \in [N]$. Hence, $\boldsymbol{S} = S_1 \times S_2 \times \cdots \times S_N$ is the set of pure strategy profiles. A pure strategy choice for all players except player i is denoted by $\boldsymbol{s}_{-i}$, i.e. $\boldsymbol{s} = (s_i, \boldsymbol{s}_{-i})$. The payoff of player i for a pure strategy profile $\boldsymbol{s} \in \boldsymbol{S}$ is denoted by $U_i(\boldsymbol{s})$ (or $U_i(s_i, \boldsymbol{s}_{-i})$), which is a random variable with probability density function (pdf) $f_i(x|\boldsymbol{s})$ and mean $u_i(\boldsymbol{s})$. The payoffs $U_i(\boldsymbol{s})$ for $i \in [N]$ and $\boldsymbol{s} \in \boldsymbol{S}$ are considered to be continuous-type random variables that are independent from each other.

Remark 5. *The same analysis holds for discrete-type random variables if the analysis is treated with a bit more subtlety as discussed in the end of this section.*

For any set S_i, let Σ_i be the set of all probability distributions over S_i. The Cartesian product of all players' mixed strategy sets, $\boldsymbol{\Sigma} = \Sigma_1 \times \Sigma_2 \times \cdots \times \Sigma_N$, is the set of mixed strategy profiles. Denote a specific mixed strategy of player i by $\sigma_i \in \Sigma_i$, where $\sigma_i(s_i)$ is the probability that player i plays strategy s_i. If the $[N] \setminus i$ players choose to play a mixed strategy $\boldsymbol{\sigma}_{-i}$, the payoff for player i if he/she plays $s_i \in S_i$ is denoted by $\overline{U}_i(s_i, \boldsymbol{\sigma}_{-i})$. Using the law of total probability, the marginal distribution of $\overline{U}_i(s_i, \boldsymbol{\sigma}_{-i})$ has the probability distribution function

$$\bar{f}_i(x|(s_i, \boldsymbol{\sigma}_{-i})) = \sum_{\boldsymbol{s}_{-i} \in \boldsymbol{S}_{-i}} \left(f_i(x|(s_i, \boldsymbol{s}_{-i})) \cdot \boldsymbol{\sigma}(\boldsymbol{s}_{-i}) \right), \tag{3.1}$$

where $\boldsymbol{\sigma}(\boldsymbol{s}_{-i}) = \prod_{j \in [N] \setminus i} \sigma_j(s_j)$ and s_j is the corresponding strategy of player j in $\boldsymbol{s}_{-i}$. Note that for $s_i \neq s_i' \in S_i$, the random variables $\overline{U}_i(s_i, \boldsymbol{\sigma}_{-i})$ and $\overline{U}_i(s_i', \boldsymbol{\sigma}_{-i})$ are not independent of each other in a single play of the game.

3.2 Risk-Averse Equilibrium

In a stochastic game where the payoffs are random variables, playing the Nash equilibrium considering the expected payoffs may create a risky situation; e.g., see [176] and [188] and the references therein for examples on multi-armed bandits. The reason is that payoffs with larger expectations may have a larger variance as well. As a result, it may be the case that playing strategies with lower expectations is more probable to have a larger payoff. This concept is mostly helpful when players play the game once, so they do not have the chance to repeat the game and gain a larger cumulative payoff by playing the strategy with the largest expected payoff. As a result, we propose the risk-averse equilibrium in a probabilistic sense rather than in an expectation sense as the Nash equilibrium. From an individual player's point of view, the best response to a mixed strategy of the rest of players is defined as follows, which is based on the notion of Risk-Averse Best Action Decision with Incomplete Information (R-ABADI).

Definition 2. *The set of mixed strategy risk-averse best responses of player i to the mixed strategy profile $\boldsymbol{\sigma}_{-i}$ is the set of all probability distributions over the set*

$$\arg\max_{s_i \in S_i} P\Big(\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\right)\Big), \tag{3.2}$$

where what we mean by $\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right)$ being greater than or equal to $\overline{\boldsymbol{U}}_i\big(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\big)$ when $S_i \setminus s_i \neq \varnothing$ is that $\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right)$ is greater than or equal to $\overline{U}_i\left(s_i', \boldsymbol{\sigma}_{-i}\right)$ for all $s_i' \in S_i \setminus s_i$; otherwise, if $S_i \setminus s_i = \varnothing$, player i only has a single option that can be played. The same randomness on the action of players $[N] \setminus i$ is considered in $\overline{U}_i(s_i, \boldsymbol{\sigma}_{-i})$ for all $s_i \in S_i$, and independent randomness on actions is discussed in Appendix B.2. We denote the risk-averse best response set of player i's strategies, given the other players' mixed strategies $\boldsymbol{\sigma}_{-i}$, by $RB(\boldsymbol{\sigma}_{-i})$, which is in general a set-valued function.

Given the definition of the risk-averse best response, the risk-averse R-ABADI equilibrium (RAE) is defined as follows. Note that the risk-averse best-response/equilibrium and R-ABADI best-response/equilibrium are used interchangeably throughout this chapter.

Definition 3. *A strategy profile* $\boldsymbol{\sigma}^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_N^*)$ *is a risk-averse R-ABADI equilibrium (RAE), if and only if* $\sigma_i^* \in RB(\boldsymbol{\sigma}_{-i}^*)$ *for all* $i \in [N]$.

The following theorem proves the existence of a mixed strategy risk-averse equilibrium for a game with finite number of players and finite number of strategies per player.

Theorem 4. *For any finite N-player game, a risk-averse equilibrium exists.*

The proof of Theorem 4 is provided in Appendix B.1.

3.2.1 Pure Strategy Risk-Averse Equilibrium

The pure strategy risk-averse best response is defined in the following as a specific case of the risk-averse best response defined in Definition 2.

Definition 4. *Pure strategy* $\widehat{s}_i$ *of player i is a risk-averse best response (RB) to the pure strategy* $\boldsymbol{s}_{-i}$ *of the other players if*

$$\begin{cases} \widehat{s}_i \in \arg\max_{s_i \in S_i} P\Big(U_i\left(s_i, \boldsymbol{s}_{-i}\right) \geq \boldsymbol{U}_i\left(S_i \setminus s_i, \boldsymbol{s}_{-i}\right)\Big), & \text{if } S_i \setminus s_i \neq \varnothing, \\ \widehat{s}_i = s_i, \quad \text{if } S_i \setminus s_i = \varnothing, \end{cases} \tag{3.3}$$

where what we mean by $U_i\left(s_i, \boldsymbol{s}_{-i}\right)$ *being greater than or equal to* $\boldsymbol{U}_i\big(S_i \setminus s_i, \boldsymbol{s}_{-i}\big)$ *is that* $U_i\left(s_i, \boldsymbol{s}_{-i}\right)$ *is greater than or equal to* $U_i\left(s_i', \boldsymbol{s}_{-i}\right)$ *for all* $s_i' \in S_i \setminus s_i$. *We denote the risk-averse best response set of player i, given the other players' pure strategies* $\boldsymbol{s}_{-i}$, *by* $RB(\boldsymbol{s}_{-i})$ *(overloading notation,* $RB(.)$ *is used for both pure and mixed strategy risk-averse best response).*

Given the definition of the pure strategy risk-averse best response, the pure strategy risk-averse equilibrium (RAE), which does not necessarily exist, is defined below.

Definition 5. *A pure strategy profile* $\boldsymbol{s}^* = (s_1^*, s_2^*, \ldots, s_N^*)$ *is a pure strategy risk-averse equilibrium (RAE), if and only if* $s_i^* \in RB(\boldsymbol{s}_{-i}^*)$ *for all* $i \in [N]$.

3.3 Strict Dominance and Iterated Elimination of Strictly Dominated Strategies

Probably the most basic solution concept for a game is the dominant strategy equilibrium. In the following definition, the strict dominance is described.

Definition 6. *A pure strategy $s_i \in S_i$ of player i strictly dominates a second pure strategy $s_i' \in S_i$ of the player if*

$$\begin{aligned} & P\Big(U_i\left(s_i, \boldsymbol{s}_{-i}\right) \geq \boldsymbol{U}_i\left(S_i \setminus s_i, \boldsymbol{s}_{-i}\right)\Big) \\ >& P\Big(U_i\left(s_i', \boldsymbol{s}_{-i}\right) \geq \boldsymbol{U}_i\left(S_i \setminus s_i', \boldsymbol{s}_{-i}\right)\Big), \forall \boldsymbol{s}_{-i} \in \boldsymbol{S}_{-i}. \end{aligned} \tag{3.4}$$

A strictly dominated strategy cannot be the risk-averse best response to any mixed strategy profile of other players due to the following reason. Consider that $s_i' \in S_i$ is strictly dominated by $s_i \in S_i$ for player i as is stated in Definition 6. Then, for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$, we have

$$\begin{aligned} & P\Big(\overline{U}_i(s_i, \boldsymbol{\sigma}_{-i}) \geq \overline{U}_i(S_i \setminus s_i, \boldsymbol{\sigma}_{-i})\Big) \\ \overset{(a)}{=} & \sum_{\boldsymbol{s}_{-i} \in \boldsymbol{S}_{-i}} \left(P\Big(U_i(s_i, \boldsymbol{s}_{-i}) \geq \boldsymbol{U}_i(S_i \setminus s_i, \boldsymbol{s}_{-i})\Big) \cdot \boldsymbol{\sigma}(\boldsymbol{s}_{-i}) \right) \\ \overset{(b)}{>} & \sum_{\boldsymbol{s}_{-i} \in \boldsymbol{S}_{-i}} \left(P\Big(U_i(s_i', \boldsymbol{s}_{-i}) \geq \boldsymbol{U}_i(S_i \setminus s_i', \boldsymbol{s}_{-i})\Big) \cdot \boldsymbol{\sigma}(\boldsymbol{s}_{-i}) \right) \\ = & P\Big(\overline{U}_i(s_i', \boldsymbol{\sigma}_{-i}) \geq \overline{U}_i(S_i \setminus s_i', \boldsymbol{\sigma}_{-i})\Big), \end{aligned} \tag{3.5}$$

where (a) is followed by using the law of total probability by partitioning on the strategies of players $[N] \setminus i$, $\boldsymbol{\sigma}(\boldsymbol{s}_{-i}) = \prod_{j \in [N] \setminus i} \sigma_j(s_j)$ and s_j is the corresponding strategy of player j in $\boldsymbol{s}_{-i}$, and (b) is true by the assumption that the pure strategy s_i' is strictly dominated by the pure strategy s_i and using Equation (3.4) in Definition 6 on strict dominance. By Equation (3.5) and Equation (3.2) in Definition 2 on the best response to a mixed strategy profile of other players, a strictly dominated pure strategy can never be a best response to any mixed strategy profile of other players. As a result, a strictly dominated pure strategy can be removed from the set of strategies of a player and iterated elimination of strictly dominated strategies can be applied to a game under the risk-averse framework.

3.4 Finding the Risk-Averse Equilibrium

The mixed strategy risk-averse equilibrium of a game can be found by choosing players' mixed strategy profiles in such a way that a player cannot strategize against other players. In other words, under a mixed strategy risk-averse equilibrium, all players are indifferent to their mixed strategies, so they use a mixed strategy to make other players indifferent as well. If all players are indifferent to their mixed risk-averse strategies, then no player has an incentive to change strategies, so we end up with a mixed strategy risk-averse equilibrium. Formally speaking, a risk-averse mixed strategy is characterized by $\sigma_i(s_i)$ for all $i \in [N]$ and for all $s_i \in S_i$, so there are $\sum_{i\in[N]} |S_i|$ parameters that should be found. Letting the mixed strategy profile for players $[N] \setminus i$ be $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$, then in order for player i to be indifferent to his/her set of strategies among a subset $S_i' \subseteq S_i$, we need to have

$$\begin{aligned} &P\Big(\overline{U}_i\left(s_i', \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i', \boldsymbol{\sigma}_{-i}\right)\Big) \\ \geq &P\Big(\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\right)\Big),\ \forall s_i \in S_i, s_i' \in S_i'. \end{aligned}$$

The above equations reveal $|S_i| - 1$ independent equations for each player i, so in total $\sum_{i\in[N]} |S_i| - N$ equations are derived. The remaining N equations are provided by the fact that the mixed strategy of each player adds to one for their set of strategies. As a result, if there is a mixed strategy risk-averse equilibrium for which only a subset $\boldsymbol{S}' = \{S_1', S_2', ..., S_N'\}$ of the pure strategies, denoted as the *support* of the equilibrium, are played with non-zero probability, this equilibrium is a solution of the following set of equations for $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$:

$$\begin{cases} \begin{aligned} &P\Big(\overline{U}_i\left(s_i', \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i', \boldsymbol{\sigma}_{-i}\right)\Big) \\ &\geq P\Big(\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\right)\Big),\ \forall s_i \in S_i, s_i' \in S_i', \forall i \in [N], \end{aligned} \\ \sum_{s_i \in S_i} \sigma_i(s_i) = 1, \forall i \in [N], \\ \sigma_i(s_i) = 0, \forall s_i \notin S_i', \forall i \in [N]. \end{cases} \tag{3.6}$$

Any solution to Equation set (3.6) is a risk-averse equilibrium, so we can check if an equilibrium exists for any support $\boldsymbol{S}' \subseteq \boldsymbol{\Sigma}$.

Note that as is stated in Equation (3.5), we have the following by using the law of total probability:

$$\begin{aligned} & P\Big(\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\right)\Big) \\ = & \sum_{s_{-i} \in \boldsymbol{S}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{s}_{-i}) \cdot P\Big(U_i(s_i, \boldsymbol{s}_{-i}) \geq \boldsymbol{U}_i(S_i \setminus s_i, \boldsymbol{s}_{-i})\Big)\right), \end{aligned} \tag{3.7}$$

where $\boldsymbol{\sigma}(\boldsymbol{s}_{-i}) = \prod_{j \in [N] \setminus i} \sigma_j(s_j)$ and s_j is the corresponding strategy of player j in $\boldsymbol{s}_{-i}$. Hence, Equation (3.7) is polynomial of order $N-1$ in terms of $\sigma(s_i)$ for $s_i \in S_i$ and $i \in [N]$. We can define a *risk-averse probability tensor* of dimension $|S_1| \times |S_2| \times \cdots \times |S_N|$, where the i-th dimension has all pure strategies $s_i \in S_i$ and each element of the tensor is an N dimensional vector defined in the following. The i-th element of the N dimensional vector corresponding to the pure strategy profile $(s_i, \boldsymbol{s}_{-i})$ is defined as

$$p_i(s_i, \boldsymbol{s}_{-i}) = P\Big(U_i\left(s_i, \boldsymbol{s}_{-i}\right) \geq \boldsymbol{U}_i\left(S_i \setminus s_i, \boldsymbol{s}_{-i}\right)\Big). \tag{3.8}$$

As a result, an equivalent approach for finding the risk-averse equilibrium is to find the Nash equilibrium of the risk-averse probability tensor, as any such Nash equilibrium must maximize the probability of playing a utility-maximizing response to $\boldsymbol{\sigma}_{-i}$ for each player i. In the following two subsections, two illustrative examples are provided to make the concept of the risk-averse equilibrium clear.

3.5 Illustrative Examples

In the following two subsections, two illustrative examples are provided to shed light on the definition of the pure and mixed strategy risk-averse equilibria.

3.5.1 Illustrative Example 3

The following example is presented to shed light on the notion of pure strategy risk-averse equilibrium.

Example 3. *Consider a game between two players where each player has two pure strategies, $S_1 = \{U, D\}$ and $S_2 = \{L, R\}$, with independent payoff distributions specified as*

(i) $U_1(U, L)$ and $U_2(U, L)$ are independent and have the same pdf as

$$f_4(u) = \alpha\Big(3e^{-20(u-2)^2} \cdot \mathbb{1}\{\frac{3}{2} \leq u \leq \frac{5}{2}\} + 2e^{-20(u-7)^2} \cdot \mathbb{1}\{\frac{13}{2} \leq u \leq \frac{15}{2}\}\Big),$$

(ii) $U_1(U, R)$ and $U_2(U, R)$ are independent and have the same pdf as

$$f_3(u) = \beta e^{-20(u-3)^2} \cdot \mathbb{1}\{\frac{5}{2} \leq u \leq \frac{7}{2}\},$$

(iii) $U_1(D, L)$ and $U_2(D, L)$ are independent and have the same pdf as

$$\widehat{f_3}(u) = \gamma\Big(3e^{-20(u-1)^2} \cdot \mathbb{1}\{\frac{1}{2} \leq u \leq \frac{3}{2}\} + 2e^{-20(u-6)^2} \cdot \mathbb{1}\{\frac{11}{2} \leq u \leq \frac{13}{2}\}\Big),$$

(iv) $U_1(D, R)$ and $U_2(D, R)$ are independent and have the same pdf as

$$f_5(u) = \delta\Big(7e^{-20(u-2)^2} \cdot \mathbb{1}\{\frac{3}{2} \leq u \leq \frac{5}{2}\} + 3e^{-20(u-12)^2} \cdot \mathbb{1}\{\frac{23}{2} \leq u \leq \frac{25}{2}\}\Big),$$

where α, β, γ, and δ are constants for which each of the corresponding distributions integrate to one and $\mathbb{1}\{.\}$ is the indicator function.

The above example is depicted in Figure 3.1. Considering the expected payoffs in Example 3 as

$$\begin{cases} \mathbb{E}\left[U_1(U, L)\right] = \mathbb{E}\left[U_2(U, L)\right] = 4, \\ \mathbb{E}\left[U_1(U, R)\right] = \mathbb{E}\left[U_2(U, R)\right] = \mathbb{E}\left[U_1(D, L)\right] = \mathbb{E}\left[U_2(D, L)\right] = 3, \\ \mathbb{E}\left[U_1(D, R)\right] = \mathbb{E}\left[U_2(D, R)\right] = 5, \end{cases}$$

the pure Nash equilibria of the game are (U, L) and (D, R), and the mixed Nash equilibrium is that the first player selects U with probability two-thirds and selects D otherwise and the second player selects L with probability two-thirds and selects R otherwise. On the other hand, it follows by using the

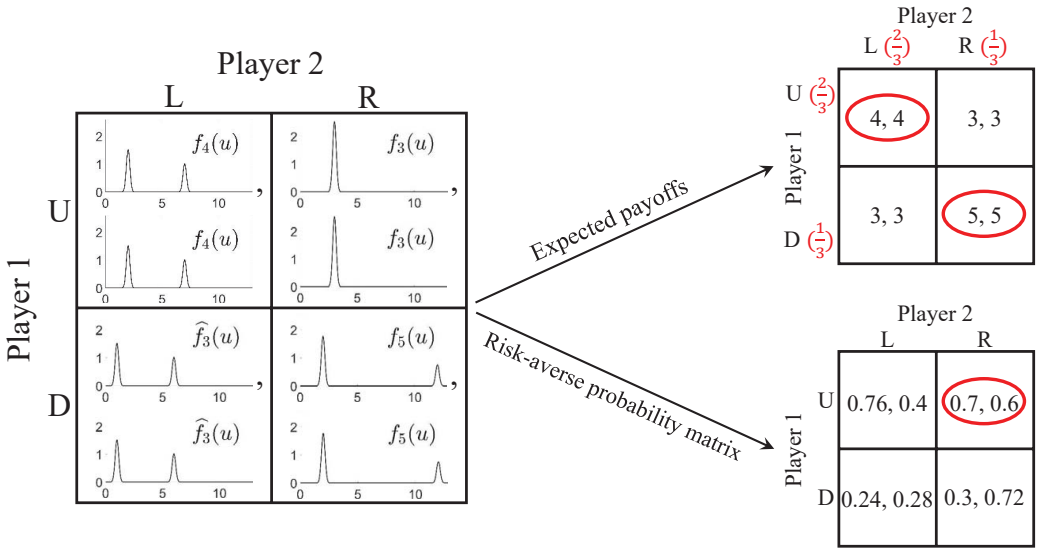

Figure 3.1: The payoff matrix of Example 3. The pure and mixed strategy Nash equilibria are shown on the top-right and the pure strategy risk-averse equilibrium is shown on the bottom-right.

payoff density functions that

$$\begin{cases} P\big(U_1(U,L) \geq U_1(D,L)\big) = 0.76, \\ P\big(U_1(U,R) \geq U_1(D,R)\big) = 0.7, \\ P\big(U_2(U,L) \geq U_2(U,R)\big) = 0.4, \\ P\big(U_2(D,L) \geq U_2(D,R)\big) = 0.28, \end{cases}$$

which are used to form the risk-averse probability bi-matrix of the game derived based on Equation (3.8). The risk-averse probability matrix is depicted in Figure 3.1. According to Definition 5, (U, R) is a pure strategy risk-averse equilibrium that is different from the Nash equilibria of the game. Taking a close look at the payoff distributions, (U, R) is less risky than (U, L) and (D, R) in a single round of the game.

3.5.2 Illustrative Example 4

In this subsection, the mixed strategy risk-averse equilibrium of a two-player game proposed in the following example is computed.

Example 4. *Consider a game between two players where each player has two pure strategies, $S_1 = \{U, D\}$ and $S_2 = \{L, R\}$, with independent payoff*

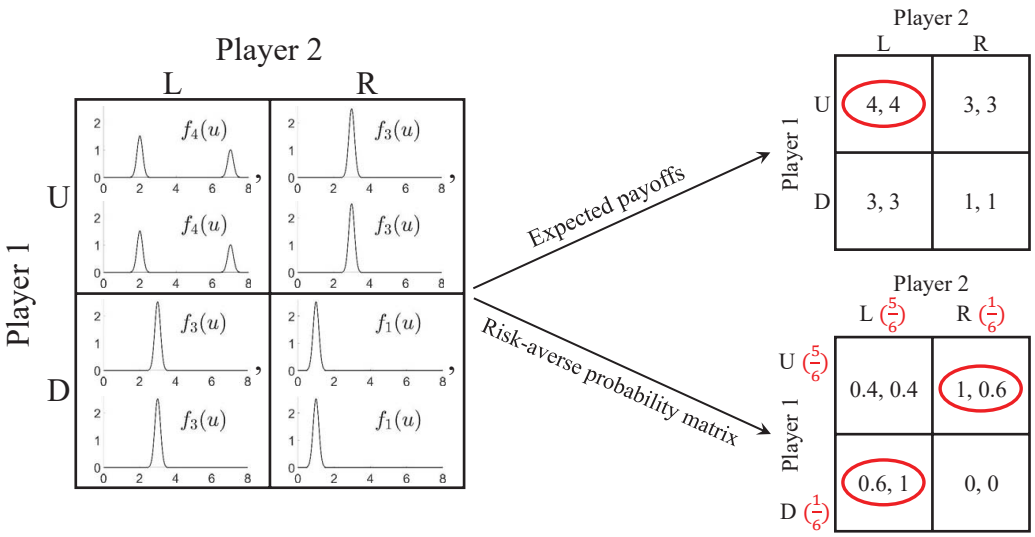

Figure 3.2: The payoff matrix of Example 4. The pure strategy Nash equilibrium is shown on the top-right and the pure and mixed strategy risk-averse equilibria are shown on the bottom-right.

distributions specified as

(i) $U_1(U,L)$ *and* $U_2(U,L)$ *are independent and have the same pdf as*

$$f_4(u) = \alpha\Big(3e^{-20(u-2)^2}\cdot\mathbb{1}\{\frac{3}{2}\le u\le\frac{5}{2}\}+2e^{-20(u-7)^2}\cdot\mathbb{1}\{\frac{13}{2}\le u\le\frac{15}{2}\}\Big),$$

(ii) $U_1(U,R), U_2(U,R), U_1(D,L),$ *and* $U_2(D,L)$ *are independent and have the same pdf as*

$$f_3(u) = \beta e^{-20(u-3)^2}\cdot\mathbb{1}\{\frac{5}{2}\le u\le\frac{7}{2}\},$$

(iii) $U_1(D,R)$ *and* $U_2(D,R)$ *are independent and have the same pdf as*

$$f_1(u) = \gamma e^{-20(u-1)^2}\cdot\mathbb{1}\{\frac{1}{2}\le u\le\frac{3}{2}\},$$

where $\alpha, \beta,$ *and* γ *are constants for which each of the corresponding distributions integrate to one.*

The above example is depicted in Figure 3.2. Considering the expected

payoffs in Example 4 as

$$\begin{cases} \mathbb{E}\left[U_1(U,L)\right] = \mathbb{E}\left[U_2(U,L)\right] = 4, \\ \mathbb{E}\left[U_1(U,R)\right] = \mathbb{E}\left[U_2(U,R)\right] = \mathbb{E}\left[U_1(D,L)\right] = \mathbb{E}\left[U_2(D,L)\right] = 3, \\ \mathbb{E}\left[U_1(D,R)\right] = \mathbb{E}\left[U_2(D,R)\right] = 1, \end{cases}$$

the pure Nash equilibrium of the game is (U, L) as depicted in Figure 3.2 with no mixed strategy Nash equilibrium. On the other hand, it follows by using the payoff density functions that

$$\begin{cases} P\big(U_1(U,L) \geq U_1(D,L)\big) = 0.4, \\ P\big(U_1(U,R) \geq U_1(D,R)\big) = 1, \\ P\big(U_2(U,L) \geq U_2(U,R)\big) = 0.4, \\ P\big(U_2(D,L) \geq U_2(D,R)\big) = 1, \end{cases}$$

which are used to form the risk-averse probability bi-matrix of the game derived based on Equation (3.8). The risk-averse probability matrix is depicted in Figure 3.2. According to Definition 5, (U, R) and (D, L) are the pure strategy risk-averse equilibria. In order to find the mixed strategy risk-averse equilibrium, consider that the first player selects U with probability σ_U and selects D otherwise. Given the first player's mixed strategy $(\sigma_U, 1-\sigma_U)$, with a little misuse of notation, denote the random variables denoting the second player's payoffs by selecting L or R with L and R, respectively. The second player is indifferent between selecting L and R if $P(L \geq R) = P(R \geq L)$. Since payoffs are continuous random variables, $P(R \geq L) = 1 - P(L \geq R)$; as a result, the second player is indifferent between the strategies if $P(L \geq R) = 0.5$. By using the law of total probability and independence of payoff distributions,

$P(L \geq R)$ can be computed as

$$\begin{aligned} P(L \geq R) &= \sigma_U \cdot P\Big(U_2(U,L) \geq U_2(U,R)\Big|\text{Player 1 plays } U\Big) \\ &\quad + (1-\sigma_U) \cdot P\Big(U_2(D,L) \geq U_2(D,R)\Big|\text{Player 1 plays } D\Big) \\ &= \sigma_U \cdot \int_{-\infty}^{\infty}\int_{v}^{\infty} f_4(u) \cdot f_3(v)\ dudv \\ &\quad + (1-\sigma_U) \cdot \int_{\infty}^{\infty}\int_{v}^{\infty} f_3(u) \cdot f_1(v)\ dudv \\ &= \frac{2}{5}\sigma_U + (1-\sigma_U) = 1 - \frac{3}{5}\sigma_U. \end{aligned} \tag{3.9}$$

Letting $P(L \geq R) = 0.5$, then $\sigma_U = \frac{5}{6}$, which determines the mixed strategy risk-averse equilibrium. As a result, due to symmetry, $\big(\sigma_1(U), \sigma_1(D)\big) = (\frac{5}{6}, \frac{1}{6})$ and $\big(\sigma_2(L), \sigma_2(R)\big) = (\frac{5}{6}, \frac{1}{6})$ form the mixed strategy risk-averse equilibrium of the game in Example 4.

It is easy to verify that the game proposed in Example 3 does not have any mixed strategy risk-averse equilibria. The game in Example 3 has both pure and mixed strategy Nash equilibria, but it only has pure strategy risk-averse equilibrium. On the other hand, the game in Example 4 only has pure strategy Nash equilibrium, but it has both pure and mixed risk-averse equilibria. As can be seen, the distributions of payoffs can have a significant impact on the behavior of players if they take risk into account when taking their decisions.

Remark 6. *As mentioned earlier in this section, the analysis for risk-averse equilibrium holds for discrete-time random variables as well. For example, consider random variables X, Y, and Z with distributions*

$$\begin{aligned} &P(X=1) = 0.8,\ P(X=2) = 0.2, \\ &P(Y=1) = 1, \\ &P(Z=1) = 0.5,\ P(Z=2) = 0.5. \end{aligned}$$

Denote the observations of the three random variables by triple (X, Y, Z) and let $\{X \geq (Y, Z)\}$ be the event that X is greater than or equal to both Y and

Z. Then

$$P\big(X \geq (Y,Z)\big) = P\big(\{(1,1,1),(2,1,1),(2,1,2)\}\big) = 0.4 + 0.1 + 0.1 = 0.6,$$
$$P\big(Y \geq (X,Z)\big) = P\big((1,1,1)\big) = 0.4,$$
$$P\big(Z \geq (X,Y)\big) = P\big(\{(1,1,1),(1,1,2),(2,1,2)\}\big) = 0.4 + 0.4 + 0.1 = 0.9.$$

As can be seen, $P\big(X \geq (Y,Z)\big)+P\big(Y \geq (X,Z)\big)+P\big(Z \geq (X,Y)\big) = 1.9 > 1$. *In order to resolve this issue, we can break ties uniformly at random as*

$$P\big(X \geq (Y,Z)\big) = \frac{1}{3} \times 0.4 + 0.1 + \frac{1}{2} \times 0.1 = \frac{17}{60},$$
$$P\big(Y \geq (X,Z)\big) = \frac{1}{3} \times 0.4 = \frac{2}{15},$$
$$P\big(Z \geq (X,Y)\big) = \frac{1}{3} \times 0.4 + 0.4 + \frac{1}{2} \times 0.1 = \frac{35}{60},$$

which results in $P\big(X \geq (Y,Z)\big) + P\big(Y \geq (X,Z)\big) + P\big(Z \geq (X,Y)\big) = 1$.

3.6 Finite-Time Commit Games

The risk-averse framework discussed in Section 3.2 provides risk-averse players with pure or mixed strategies such that given the other players' strategies, risk-averse equilibrium maximizes the probability that a player is rewarded the most in a single round of the game rather than maximizing the expected received reward. On the other hand, for infinite rounds of playing the game, given the other players' strategies, selecting the strategy that maximizes the expected reward guarantees maximum cumulative reward. However, the rewards may not be satisfying for a risk-averse player in each and every round of playing the game. As a result, risk-averse players may even choose to play the risk-averse equilibrium in infinite (or finite) rounds of games to have more or less balanced rewards in all rounds of the game rather than have maximum cumulative reward in the end. Despite this fact, we present a slightly different approach for finite-time games that aims to maximize not the expected cumulative reward but rather the probability of receiving the highest cumulative reward. Note that the proposed equilibrium for finite-time commit games in this section may be different from the Nash equilibrium or the equilibrium presented in Section 3.2.

Consider that the N players plan to play a game for M independent times where all players have to commit to the pure strategy they play in the first round for the whole game. The strategy in the first round of the game does not have to be pure and can be mixed, but a player has to commit to the randomly sampled pure strategy according to the mixed strategy for M times. Let $U_i^M(s_i, \boldsymbol{s}_{-i}) = U_i^M(\boldsymbol{s}) = X^1 + X^2 + \cdots + X^M$, where X^j for $1 \leq j \leq M$ are independent and identically distributed random variables and $X^1 \sim f_i(x|\boldsymbol{s})$. If players choose to play $\boldsymbol{s} \in \boldsymbol{S}$ for the whole game with M rounds, the random variable $U_i^M(\boldsymbol{s})$ denotes the cumulative payoff for player $i \in [N]$ in the end of the M plays and $U_i^M(\boldsymbol{s}) \sim f_i^M(x|\boldsymbol{s}) = \underbrace{f_i(x|\boldsymbol{s}) \circledast \cdots \circledast f_i(x|\boldsymbol{s})}_{M \text{ times}}$.

If the $[N] \setminus i$ players choose to play a mixed strategy $\boldsymbol{\sigma}_{-i}$ in the first round of the game and commit to it for $M-1$ other rounds of the game, using the law of total probability, the distribution of the cumulative payoff for player i in the end of the game when he/she plays s_i, denoted by $\overline{U}_i^M(s_i, \boldsymbol{\sigma}_{-i})$, has the probability distribution function

$$\bar{f}_i^M(x|(s_i, \boldsymbol{\sigma}_{-i})) = \sum_{\boldsymbol{s}_{-i} \in \boldsymbol{S}_{-i}} \left(f_i^M(x|(s_i, \boldsymbol{s}_{-i})) \cdot \boldsymbol{\sigma}(\boldsymbol{s}_{-i}) \right), \tag{3.10}$$

where $\boldsymbol{\sigma}(\boldsymbol{s}_{-i}) = \prod_{j \in [N] \setminus i} \sigma_j(s_j)$ and s_j is the corresponding strategy of player j in $\boldsymbol{s}_{-i}$. Note that for $s_i, s_i' \in S_i$, the random variables $\overline{U}_i^M(s_i, \boldsymbol{\sigma}_{-i})$ and $\overline{U}_i^M(s_i', \boldsymbol{\sigma}_{-i})$ are not independent from each other in a single play of the game. The risk-averse equilibrium for an M-time commit game can be derived similarly to the derivations in Section 3.2 and is described below. From an individual player's point of view, the best response to a mixed strategy of the rest of the players for an M-time commit game is defined as follows.

Definition 7. *The set of mixed strategy risk-averse best responses of player i to the mixed strategy profile $\boldsymbol{\sigma}_{-i}$ for an M-time commit game is the set of all probability distributions over the set*

$$\arg\max_{s_i \in S_i} P\Big(\overline{U}_i^M\left(s_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i^M\left(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\right) \Big), \tag{3.11}$$

where what we mean by $\overline{U}_i^M\left(s_i, \boldsymbol{\sigma}_{-i}\right)$ being greater than or equal to $\overline{\boldsymbol{U}}_i^M\big(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\big)$ is that $\overline{U}_i^M\left(s_i, \boldsymbol{\sigma}_{-i}\right)$ is greater than or equal to $\overline{U}_i^M\left(s_i', \boldsymbol{\sigma}_{-i}\right)$ for all $s_i' \in S_i \setminus s_i$; otherwise, if $S_i \setminus s_i = \varnothing$, player i has only a single option to play.

We denote the risk-averse best response set of player i's mixed strategies for an M-time commit game, given the other players' mixed strategies $\boldsymbol{\sigma}_{-i}$, by $RB^M(\boldsymbol{\sigma}_{-i})$, which is a set-valued function.

Given the definition of the risk-averse best response for M-time commit games, the risk-averse equilibrium (RAE) for M-time commit games is defined as follows.

Definition 8. *A strategy profile $\boldsymbol{\sigma}^{*M} = (\sigma_1^{*M}, \sigma_2^{*M}, \ldots, \sigma_N^{*M})$ is a risk-averse equilibrium (RAE) for an M-time commit game, if and only if $\sigma_i^{*M} \in RB^M(\boldsymbol{\sigma}_{-i}^{*M})$ for all $i \in [N]$.*

The following corollary is resulted directly from Theorem 4.

Corollary 5. *For any finite N-player finite-time commit game, a risk-averse equilibrium exists.*

The pure strategy risk-averse best response for an M-time commit game is defined in the following as a specific case of the risk-averse best response defined in Definition 7.

Definition 9. *Pure strategy $\widehat{s}_i$ of player i is a risk-averse best response (RB) to the pure strategy $\boldsymbol{s}_{-i}$ of the other players for an M-time commit game if*

$$\begin{cases} \widehat{s}_i \in \arg\max_{s_i \in S_i} P\Big(U_i^M\left(s_i, \boldsymbol{s}_{-i}\right) \geq \boldsymbol{U}_i^M\left(S_i \setminus s_i, \boldsymbol{s}_{-i}\right)\Big), & \text{if } S_i \setminus s_i \neq \varnothing, \\ \widehat{s}_i = s_i, \quad \text{if } S_i \setminus s_i = \varnothing, \end{cases} \tag{3.12}$$

where what we mean by $U_i^M\left(s_i, \boldsymbol{s}_{-i}\right)$ being greater than or equal to $\boldsymbol{U}_i^M\big(S_i \setminus s_i, \boldsymbol{s}_{-i}\big)$ is that $U_i^M\left(s_i, \boldsymbol{s}_{-i}\right)$ is greater than or equal to $U_i^M\left(s_i', \boldsymbol{s}_{-i}\right)$ for all $s_i' \in S_i \setminus s_i$. We denote the risk-averse best response set of player i for an M-time commit game, given the other players' pure strategies $\boldsymbol{s}_{-i}$, by $RB^M(\boldsymbol{s}_{-i})$ (overloading notation, $BR^M(.)$ is used for both mixed and pure strategy risk-averse best response for M-time commit games).

Given the definition of the pure strategy risk-averse best response for an M-time commit game, the pure strategy risk-averse equilibrium (RAE), which does not necessarily exist, is defined below.

Definition 10. *A pure strategy profile $\boldsymbol{s}^{*M} = (s_1^{*M}, s_2^{*M}, \ldots, s_N^{*M})$ is a pure strategy risk-averse equilibrium (RAE) for an M-time commit game, if and only if $s_i^{*M} \in RB^M(\boldsymbol{s}_{-i}^{*M})$ for all $i \in [N]$.*

3.7 Numerical Results

In this section, the classical Nash equilibrium is compared with the proposed risk-averse equilibrium. To this end, the likelihood of receiving the higher reward in a two-player game is evaluated under the two types of equilibria for the following example.

Example 5. *Consider a game between two players where each player has two pure strategies, $S_1 = \{U, D\}$ and $S_2 = \{L, R\}$, with independent payoff distributions specified as*

(i) $U_1(U, L)$ and $U_2(U, L)$ are independent and have the same pdf as

$$\begin{aligned} f_{1,1}(u) =& \alpha\Big(3e^{-20(u-1)^2} \cdot \mathbb{1}\{\frac{1}{2} \leq u \leq \frac{3}{2}\} \\ &+ 2e^{-20(u-a)^2} \cdot \mathbb{1}\{a - \frac{1}{2} \leq u \leq a + \frac{1}{2}\}\Big), \end{aligned}$$

(ii) $U_1(U, R)$, $U_2(U, R)$, $U_1(D, L)$, and $U_2(D, L)$ are independent and have the same pdf as

$$f_{1,2}(u) = \beta e^{-20(u-3)^2} \cdot \mathbb{1}\{\frac{5}{2} \leq u \leq \frac{7}{2}\},$$

(iii) $U_1(D, R)$ and $U_2(D, R)$ are independent and have the same pdf as

$$\begin{aligned} f_{2,2}(u) =& \gamma\Big(7e^{-20(u-2)^2} \cdot \mathbb{1}\{\frac{3}{2} \leq u \leq \frac{5}{2}\} \\ &+ 3e^{-20(u-a-2)^2} \cdot \mathbb{1}\{a + \frac{3}{2} \leq u \leq a + \frac{5}{2}\}\Big), \end{aligned}$$

where α, β, and γ are constants for which each of the corresponding distributions integrate to one, $a \geq 0$ is a constant, and $\mathbb{1}\{.\}$ is the indicator function.

The Nash equilibrium in the above example depends on the value of the constant a. If $0 \leq a < 3.333$, there are two pure Nash equilibria (U, R) and (D, L) for the game in addition to a mixed Nash equilibrium. If $3.333 \leq a \leq 6$, the pure strategy (D, R) is the only Nash equilibrium of the game. If $a > 6$, pure strategies (U, L) and (D, R) are the two pure Nash equilibria of the game in addition to a mixed Nash equilibrium. On the other hand, no matter what the value of the constant a is, the game has a mixed risk-averse equilibrium as well as two pure risk-averse equilibria, which are (U, R) and

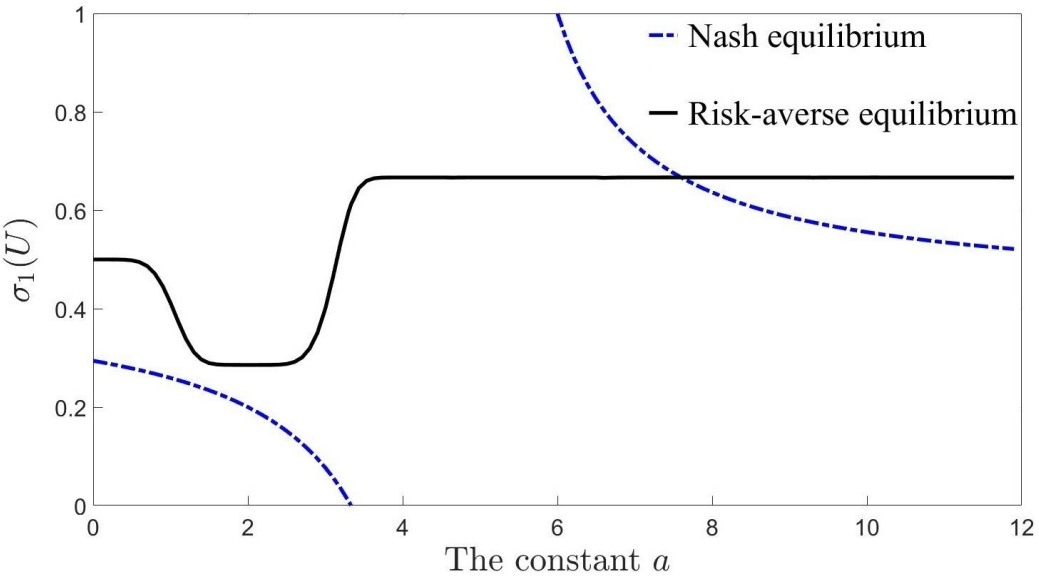

Figure 3.3: The mixed strategy Nash and risk-averse equilibria are determined by the value of $\sigma_1(U)$ in Example 5. The mixed strategies depend on the value of the constant a, where $\sigma_1(U)$ is plotted above as a function of the constant a.

(D, L). The mixed strategy Nash equilibrium and the mixed strategy risk-averse equilibrium depend on the value of the constant a. Note that the game is symmetric from the perspective of the two players, so the mixed strategy Nash and risk-averse equilibria are characterized by $\sigma_1(U)$ that is plotted in Figure 3.3.

A game according to Example 5 is simulated for 10^6 rounds for a fixed constant a. In each realization of the game, both Nash and risk-averse equilibria are played and their corresponding payoffs are compared for one of the players to see which one is larger. The mixed strategies under Nash and risk-averse equilibria are compared against each other and the pure strategies under Nash and risk-averse equilibria, if different, are compared as well. After the 10^6 games, the proportion of the games in which playing according to the risk-averse equilibrium outperforms playing according to the Nash equilibrium by having a larger payoff is computed and plotted in Figure 3.4 as a function of the constant a. In the plot in Figure 3.4, the curves comparing the Nash and risk-averse mixed strategies are dotted lines, the curve comparing the Nash equilibrium (D, R) and the risk-averse pure strategy (U, R) (or (D, L)) is a solid line for $a > 3.333$, and the curve comparing the Nash equilibrium (U, L) and risk-averse pure strategy (U, R) (or (D, L)) is a dash-

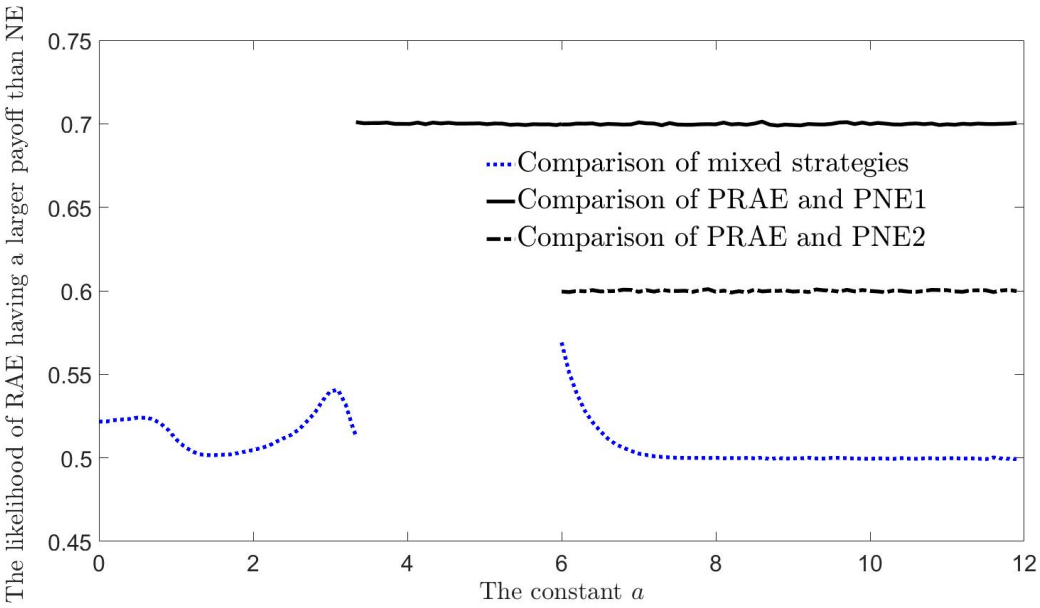

Figure 3.4: The likelihood of the payoff of the risk-averse equilibrium being greater than the payoff of the Nash equilibrium.

dotted line for $a > 6$. Note that the payoff distributions are the same for pure risk-averse equilibria (PRAE) (U, R) and (D, L), and that is why pure Nash equilibria (D, R) (PNE1) and (U, L) (PNE2) are compared with only one of the PRAEs. For the interval $0 \leq a < 3.333$ the pure equilibria for both risk-averse and Nash are the same under both approaches so neither of the pure equilibria comparison curves are shown. Figure 3.4 demonstrates that agents who play in a risk-averse manner are more likely to receive a higher payoff than under the Nash equilibrium in any single play of the game.

Chapter 4

RISK-AVERSE EQUILIBRIUM FOR AUTONOMOUS VEHICLES IN STOCHASTIC CONGESTION GAMES

The intelligent transportation systems are growing faster than ever with the speedy emergence of autonomous vehicles, unmanned aerial vehicles, Amazon delivery robots, Uber/Lyft self-driving cars, and such. One of the principal components of such systems is the navigation system whose goal is to provide travelers with fast and reliable paths from their sources to destinations. In a fleet of vehicles, an equilibrium is achieved when no travelers have any incentives in a certain sense to change routes unilaterally. In the classical Wardrop equilibrium [82, 189], travelers have incentives to change routes if they have an alternative route that has lower expected travel time. In other words, the optimality metric is based on minimizing the expected travel time in the Wardrop equilibrium. In the context of transportation though, collisions, weather conditions, road work, traffic signals, and varying traffic conditions can cause deviations in travel times [88]. As a result, the path with the minimum expected travel time may not be reliable due to its high variability. Similarly, in the context of telecommunication networks, noise, signal degradation, interference, re-transmission, and malfunctioning equipment can cause variability in transmission time from source to destination [88]. The empirical works by Abdel-Aty et al. [190], Kazimi et al. [191], Lam [192], Lam and Small [193], and Small [194] also support the fact that taking travel time uncertainty into account is indeed an essential criterion in navigation systems.

As mentioned above, minimizing the expected travel time is inadequate in scenarios involving risk due to variability of travel times. In order to address this issue, we study a richer class of congestion games called stochastic congestion games in an atomic setting, where the travel times along different arcs of the network are random variables that are not necessarily independent of each other and atomic games are those with finite numbers of players. In this framework, we introduce probability statements regarding the Risk-

Averse Best Action Decision with Incomplete Information (R-ABADI) of a traveler given the choice of the rest of travelers in the network. We propose three classes of risk-averse equilibria for stochastic congestion games: risk-averse R-ABADI equilibrium (RAE), mean-variance equilibrium (MVE), and conditional value at risk level α equilibrium (CVaR$_\alpha$E), whose notions of risk-averse best responses are based on maximizing the probability of traveling along the shortest path (also known as Risk-Averse Best Action Decision with Incomplete Information (R-ABADI)), minimizing a linear combination of mean and variance of path delay, and minimizing the expected delay at a specified risky quantile of the delay distributions, respectively. We prove that the R-ABADI, mean-variance, and CVaR$_\alpha$ equilibria exist for any finite stochastic atomic congestion game. Note that two equilibria similar to the mean-variance and CVaR equilibria exist in the literature and are discussed in the related work section, but the probability distributions of travel times are load independent or link delays are considered to be independent in the literature, which is not the case in this work. It is noteworthy that most studies on stochastic congestion games make use of simplifying assumptions such as considering the arc delay distributions to be independent of their loads or adding independent and identically distributed errors to nominal delays of arcs neglecting their differences. In the Braess paradox [195], [196], which is known to be a counterintuitive example rather than a paradox, the risk-neutral/selfish travelers select the shortest path in expected travel time, which maximizes the social delay/cost incurred by the whole society. Although the focus of this chapter is not on deriving bounds on price of anarchy, we study the Braess paradox in a stochastic setting under the three proposed risk-averse equilibria and show that the risk-averse behavior of travelers results in improving the social delay/cost incurred by the society; and as a result, the price of anarchy is improved if travelers are risk-averse. As the result, the Braess paradox may not occur to the extent presented originally if travelers are risk-averse. Furthermore, we study the Pigou network [197] in a stochastic setting and observe that the price of anarchy is also improved if travelers are risk-averse in the senses discussed above. Note that the Pigou networks are prevalent in traffic/telecommunication networks. Hence, providing travelers with risk-averse navigation can decrease the social delay/cost in the real world applications.

The rest of the chapter is structured in the following way. The stochastic

congestion game is formally defined in Section 4.1. The three proposed classes of equilibria, i.e. risk-averse R-ABADI, mean-variance, and CVaR_α equilibria, are presented in Section 4.2 and their existences in any finite stochastic congestion game are proven; detailed proofs can be found in Appendix C. Numerical results including the study of the Pigou and Braess networks as well as notes for practitioners are provided in Section 4.3. Conclusions and discussion of opportunities for future work are provided in Chapter 6.

4.1 Problem Statement of Stochastic Congestion Games

Consider a directed graph (network) $G = (\mathcal{N}, \mathcal{E})$ with a node set $\mathcal{N} = [N] \coloneqq \{1, 2, \ldots, N\}$ and directed link (edge) set $\mathcal{E}$ with cardinality $|\mathcal{E}|$, where the pair $(i, j) \in \mathcal{E}$ indicates a directed link from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$ in the directed graph. Denote the set of source-destination (SD) pairs with $\mathcal{K} \subseteq \mathcal{N} \times \mathcal{N}$, where for the SD pair $k = (s_k, d_k) \in \mathcal{K}$, $s_k \neq d_k$, the set of simple directed paths from s_k to d_k in G is denoted by $\mathcal{P}_k$, and let n_k be the number of players (travelers, vehicles, or data packages) associated with source-destination k. Let $\mathcal{P} \coloneqq \cup_{k \in \mathcal{K}} \mathcal{P}_k$ be the set of all paths. A feasible assignment $\boldsymbol{m} \coloneqq \{m^p : p \in \mathcal{P}\}$ allocates a non-negative number of players to every path $p \in \mathcal{P}$ such that $\sum_{p \in \mathcal{P}_k} m^p = n_k$ for all $k \in \mathcal{K}$. As a result, the number of players along link $e \in \mathcal{E}$ denoted by m_e is given by $m_e = \sum_{\{p \in \mathcal{P} : e \in p\}} m^p$.

The latency (delay or travel time) along link e is load-dependent which is denoted by the non-negative continuous random variable $L_e(m_e)$ with marginal probability density function (pdf) $f_e(x|m_e)$ and mean $l_e(m_e)$. Note that the number of players along an edge is determined by an assignment $\boldsymbol{m}$, so $L_e(\boldsymbol{m})$, $f_e(x|\boldsymbol{m})$, and $l_e(\boldsymbol{m})$ can be used instead of $L_e(m_e)$, $f_e(x|m_e)$, and $l_e(m_e)$, respectively. Furthermore, the latency along links of the graph can be dependent, in which case, the joint pdf of latency over all links is denoted by $f_{e_1, e_2, \ldots, e_{|\mathcal{E}|}}(x_1, x_2, \ldots, x_{|\mathcal{E}|}|m_1, m_2, \ldots, m_{|\mathcal{E}|})$, which can be denoted as $f_{\mathcal{E}}(x_1, x_2, \ldots, x_{|\mathcal{E}|}|\boldsymbol{m})$. Given the link latency defined above, the nominal latency of player i along path $p_i \in \mathcal{P}$ under a given assignment $\boldsymbol{m}$ is simply

$L^i(\boldsymbol{m}) \coloneqq \sum_{e\in p_i} L_e(\boldsymbol{m})$ with pdf

$$f^i(x|\boldsymbol{m}) = \partial \left(\int \int \cdots \int_{\{\sum_{e\in p_i} x_e \le x\}} f_{\mathcal{E}}(x_1, x_2, \ldots, x_{|\mathcal{E}|}|\boldsymbol{m})\, dx_1 dx_2 \ldots dx_{|\mathcal{E}|} \right) \Big/ \partial x$$

and mean $l^i(\boldsymbol{m}) = \sum_{e\in p_i} l_e(\boldsymbol{m})$.

The stochastic congestion game consists of $n \coloneqq \sum_{k\in\mathcal{K}} n_k$ players (travelers), where player $i \in [n] \coloneqq \{1, 2, \ldots, n\}$ is associated with the corresponding source-destination pair $k(i) \in \mathcal{K}$. As a result, $\mathcal{P}_{k(i)}$ is the set of possible pure strategies (actions) for player i. The pure strategy profile of all n players is denoted by $\boldsymbol{p} \coloneqq (p_1, p_2, \ldots, p_n)$, where $p_i \in \mathcal{P}_{k(i)}$, that fully specifies all actions in the game. The set of all pure strategy profiles is the Cartesian product of pure strategy sets of all players which is denoted by $\boldsymbol{\mathcal{P}} \coloneqq \mathcal{P}_{k(1)} \times \mathcal{P}_{k(2)} \cdots \times \mathcal{P}_{k(n)}$. Let $\boldsymbol{p}_{-i} \coloneqq (p_1, p_2, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n)$ be the pure strategies of all players except player i, so $\boldsymbol{p} = (p_i, \boldsymbol{p}_{-i})$. Given the pure strategy profile $\boldsymbol{p}$, the number of players on a path $p \in \mathcal{P}$ is given by $m^p = \sum_{i=1}^n \mathbb{1}\{p_i = p\}$, and the number of players on a link $e \in \mathcal{E}$ is given by $m_e = \sum_{\{p\in\mathcal{P}: e\in p\}} \sum_{i=1}^n \mathbb{1}\{p_i = p\}$. Let $\boldsymbol{m}(\boldsymbol{p})$ show the number of players on all paths which is fully determined by the pure strategy $\boldsymbol{p}$. As a result, given the pure strategy profile $\boldsymbol{p} = (p_i, \boldsymbol{p}_{-i})$, the latency of player i by choosing the path p_i is the random variable $L^i(\boldsymbol{m}(\boldsymbol{p})) = \sum_{e\in p_i} L_e(\boldsymbol{m}(\boldsymbol{p}))$ with pdf $f^i(x|\boldsymbol{m}(\boldsymbol{p}))$ and mean $l^i(\boldsymbol{m}(\boldsymbol{p})) = \sum_{e\in p_i} l_e(\boldsymbol{m}(\boldsymbol{p}))$. For simplicity, instead of using $L^i(\boldsymbol{m}(\boldsymbol{p}))$, $f^i(x|\boldsymbol{m}(\boldsymbol{p}))$, and $l^i(\boldsymbol{m}(\boldsymbol{p}))$, we use $L^i(\boldsymbol{p})$, $f^i(x|\boldsymbol{p})$, and $l^i(\boldsymbol{p})$, respectively.

The mixed strategy of player i is denoted by $\sigma_i \in \Sigma_i$, where Σ_i is the set of all probability distributions over the set of pure strategies $\mathcal{P}_{k(i)}$, and $\sigma_i(p)$ is the probability that player i selects path p. The mixed strategy profile of all n players is denoted by $\boldsymbol{\sigma} \coloneqq (\sigma_1, \sigma_2, \ldots, \sigma_n)$, where $\sigma_i \in \Sigma_i$. The set of all mixed strategy profiles is the Cartesian product of mixed strategy sets of all players which is denoted by $\boldsymbol{\Sigma} \coloneqq \Sigma_1 \times \Sigma_2 \cdots \times \Sigma_n$. Let $\boldsymbol{\sigma}_{-i} \coloneqq (\sigma_1, \sigma_2, \ldots, \sigma_{i-1}, \sigma_{i+1}, \ldots, \sigma_n)$ be the mixed strategies of all players except player i, so $\boldsymbol{\sigma} = (\sigma_i, \boldsymbol{\sigma}_{-i})$. The latency of player i by selecting path p_i when the other $[n] \setminus i$ players select paths according to a mixed strategy $\boldsymbol{\sigma}_{-i}$ is denoted by the random variable $\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})$ that has the following pdf

using the law of total probability:

$$\bar{f}^i(x|(p_i, \boldsymbol{\sigma}_{-i})) = \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \Big(f^i(x|(p_i, \boldsymbol{p}_{-i})) \cdot \boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \Big), \tag{4.1}$$

where $\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) = \prod_{j \in [n] \setminus i} \sigma_j(p_j)$ and p_j is the corresponding strategy of player j in $\boldsymbol{p}_{-i}$, and the mean of the random variable is given as

$$\bar{l}^i(p_i, \boldsymbol{\sigma}_{-i}) \coloneqq \mathbb{E}[\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})] = \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \Big(l^i(p_i, \boldsymbol{p}_{-i}) \cdot \boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \Big). \tag{4.2}$$

The expected average delay (latency) incurred by the n players in the stochastic congestion game under the pure strategy profile $\boldsymbol{p}$, also known as the *social cost* or *social delay* in this context, is denoted by $D(\boldsymbol{p}) \coloneqq \frac{1}{n} \sum_{i=1}^{n} l^i(\boldsymbol{p})$. The social delay under the mixed strategy $\boldsymbol{\sigma}$ is $D(\boldsymbol{\sigma}) \coloneqq \frac{1}{n} \sum_{\boldsymbol{p} \in \boldsymbol{\mathcal{P}}} \sum_{i=1}^{n} \boldsymbol{\sigma}(\boldsymbol{p}) \cdot l^i(\boldsymbol{p})$, where $\boldsymbol{\sigma}(\boldsymbol{p}) = \prod_{i \in [n]} \sigma_i(p_i)$ and p_i is the corresponding strategy of player i in $\boldsymbol{p}$. The (pure) optimal load assignment denoted by $\boldsymbol{o}$ minimizes social delay among all possible (pure) load assignments which might be in contrast with the selfish behavior of players. The (pure) *price of anarchy* (PoA) of a congestion game is the maximum ratio $D(\boldsymbol{p})/D(\boldsymbol{o})$ over all equilibria $\boldsymbol{p}$ of the game. Throughout the chapter, we follow the convention that $y \leq \boldsymbol{x}$ means that y is less than or equal to all elements of the vector $\boldsymbol{x}$.

4.2 Risk-Averse Equilibrium for Stochastic Congestion Games

In the following subsection, illustrative examples are provided with analysis of their equilibria in classic and risk-averse frameworks which motivate the novel risk-averse best-response approach for incomplete information congestion games presented in this chapter.

4.2.1 Illustrative Examples

The Pigou network [198] is one of the simplest networks studied in congestion games. We first use the Pigou network to clearly state the motivation

of the current work in the first example. We then study the more controversial network used by Braess [195] in the famous Braess's paradox in the second example. The two examples below set grounding for the risk-averse equilibrium for congestion games proposed in this chapter.

Example 6. *Consider the Pigou network with two parallel links between source and destination as shown in Figure 4.1. There are n players (vehicles or data packages) to travel from source to destination. The top and bottom links are labeled as* 1 *and* 2 *with loads* m_1 *and* $m_2 = n - m_1$*, respectively. The travel times on links* 1 *and* 2 *are respectively independent random variables* $L_1(m_1)$ *and* $L_2(m_2)$ *with expected values* $l_1(m_1) = \frac{m_1}{n}$ *and* $l_2(m_2) = 1$ *and pdfs*

$$f_1(x|m_1) = \alpha\Bigg(2\exp\Big(-100\Big(x-\frac{m_1}{4n}\Big)^2\Big)\cdot\mathbb{1}\Big\{0\le x\le\frac{m_1}{2n}\Big\} + 3\exp\Big(-100\Big(x-\frac{3m_1}{2n}\Big)^2\Big)\cdot\mathbb{1}\Big\{\frac{5m_1}{4n}\le x\le\frac{7m_1}{4n}\Big\}\Bigg),$$

$$f_2(x|m_2) = \beta\exp\left(-100\,(x-1)^2\right)\cdot\mathbb{1}\left\{\frac{3}{4}\le x\le\frac{5}{4}\right\},$$

where α *and* β *are constants for which each of the two distributions integrate to one and* $\mathbb{1}\{.\}$ *is the indicator function.*

The well-known Wardrop equilibrium [82,189], also Nash equilibrium [28], for the Pigou network in Example 6 is that all the n players travel along the top link since it is the weakly dominant strategy for any player as the expected latency incurred along the top link is always less than or equal to the expected latency incurred along the bottom link, $l_1(m_1) = \frac{m_1}{n} \le 1 = l_2(m_2)$. As a result, the Wardrop equilibrium for Pigou network is $\boldsymbol{p}^*_W = (1, 1, \ldots, 1)$ with social delay $D_W(\boldsymbol{p}^*_W) = 1$. However, although the expected latency along the top link is less than or equal to that of the bottom link, $l_1(m_1) \le l_2(m_2)$, the variance of travel time along the top link at full capacity is larger than that along the bottom link, which increases the risk and uncertainty of traveling along the top link. In fact, the bottom link with higher expected travel time is more likely to have a lower delay than the top link at full capacity; i.e., $P\big(L_2(0) \le L_1(n)\big) = 0.6 > 0.5$. As a result, a risk-averse player selects the bottom link for commute when the top link is at full capacity, especially if it is a one-time trip, for which, as is shown later, the risk-averse behavior of

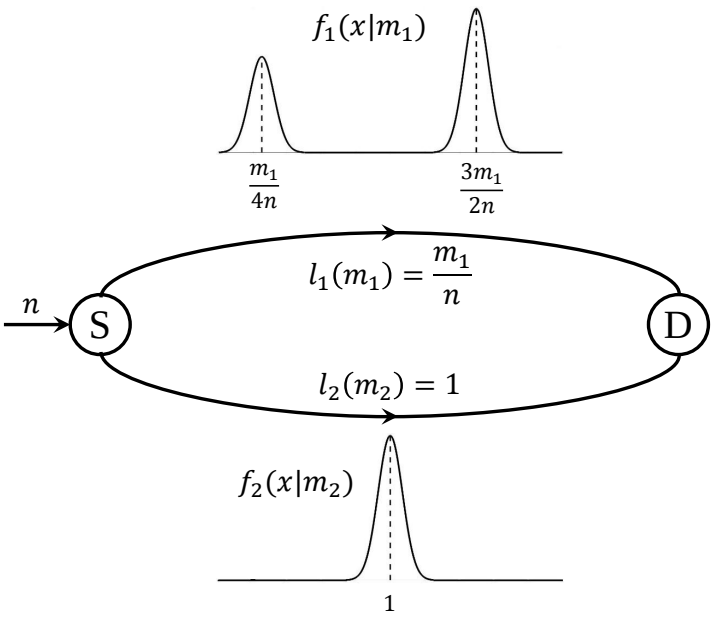

Figure 4.1: The Pigou network in Example 6 with the load-dependent latency pdfs and the corresponding means of links.

players decreases social delay for this example. As an example, consider a traveler who wants to go from hotel to airport who has two options for this trip: taking the highway that has lower expected travel time, but is more likely to get congested due to traffic jams and crashes (top link in Pigou network), or taking the urban streets with a higher expected travel time and lower congestion (the bottom link in Pigou network). A risk-neutral player travels along the top link with lower expected latency, but a risk-averse player travels along the bottom link to assure not to incur a long delay and miss the flight. Even in everyday commutes between home and work, the expected delay over many days may not be a desirable objective to minimize. No-one desires to arrive early to work some days but late on others, and to be penalized accordingly. The Braess network, studied in the next example, enforces the fact that minimizing the expected delay is not desirable for risk-averse players.

Example 7. *Consider the Braess network depicted in Figure 4.2. There are n players (vehicles or data packages) to travel from source to destination. Other than the source and destination, there are two nodes A and B in the network. The directed links (S, A), (A, D), (S, B), (B, D), and (A, B) are referred to as links 1, 2, 3, 4, and 5 with loads m_1, m_2, m_3, m_4, and m_5,*

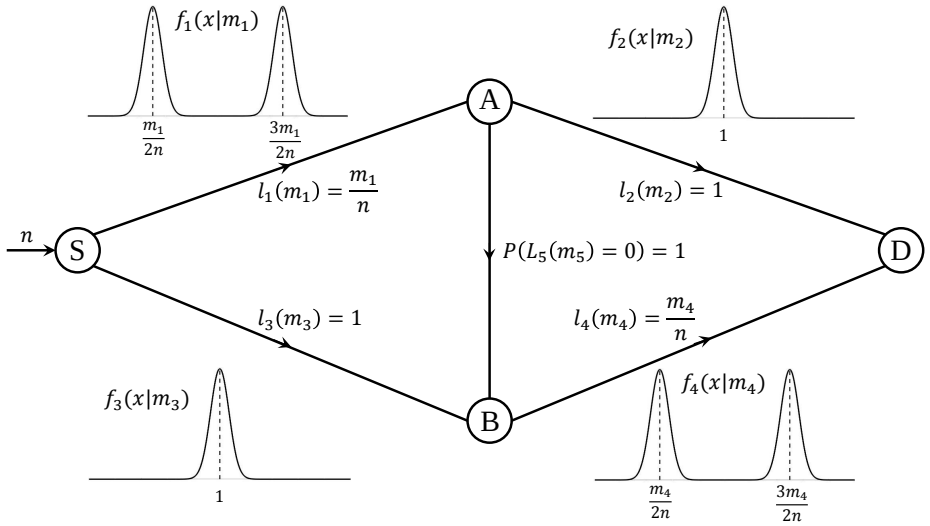

Figure 4.2: The Braess network in Example 7 with the load-dependent latency pdfs and the corresponding means of links.

respectively. The travel times on links 1, 2, 3, 4, *and* 5 *are respectively independent random variables* $L_1(m_1)$, $L_2(m_2)$, $L_3(m_3)$, $L_4(m_4)$, *and* $L_5(m_5)$ *with expected values* $l_1(m_1) = \frac{m_1}{n}$, $l_2(m_2) = 1$, $l_3(m_3) = 1$, $l_4(m_4) = \frac{m_4}{n}$, *and* $l_5(m_5) = 0$ *and pdfs*

$$
\begin{aligned}
f_1(x|m_1) &= \gamma\Bigg(\exp\Bigg(-100\Big(x - \frac{m_1}{2n}\Big)^2\Bigg) \cdot \mathbb{1}\left\{0 \le x \le \frac{m_1}{n}\right\} \\
&\qquad + \exp\Bigg(-100\Big(x - \frac{3m_1}{2n}\Big)^2\Bigg) \cdot \mathbb{1}\left\{\frac{m_1}{n} < x \le \frac{2m_1}{n}\right\}\Bigg), \\
f_2(x|m_2) &= \zeta \exp\left(-100\,(x-1)^2\right) \cdot \mathbb{1}\left\{\frac{1}{2} \le x \le \frac{3}{2}\right\}, \\
f_3(x|m_3) &= \zeta \exp\left(-100\,(x-1)^2\right) \cdot \mathbb{1}\left\{\frac{1}{2} \le x \le \frac{3}{2}\right\}, \\
f_4(x|m_4) &= \gamma\Bigg(\exp\Bigg(-100\Big(x - \frac{m_4}{2n}\Big)^2\Bigg) \cdot \mathbb{1}\left\{0 \le x \le \frac{m_4}{n}\right\} \\
&\qquad + \exp\Bigg(-100\Big(x - \frac{3m_4}{2n}\Big)^2\Bigg) \cdot \mathbb{1}\left\{\frac{m_4}{n} < x \le \frac{2m_4}{n}\right\}\Bigg),
\end{aligned}
$$

where γ *and* ζ *are constants for which the distributions integrate to one,* $\mathbb{1}\{.\}$ *is the indicator function, and* $P\big(L_5(m_5) = 0\big) = 1$. *There are three paths from source to destination,* (S, A, D), (S, A, B, D), *and* (S, B, D), *that are referred*

to as paths $1, 2$, *and* 3 *with loads* m^1, m^2, *and* m^3, *respectively, where the difference between links and paths should be clear from the context. Note that the link loads are related to path loads as* $m_1 = m^1 + m^2$, $m_2 = m^1$, $m_3 = m^3$, $m_4 = m^2 + m^3$, *and* $m_5 = m^2$, *and the delays along paths are related to link delays as* $L^1(\boldsymbol{m}) = L_1(m_1) + L_2(m_2)$, $L^2(\boldsymbol{m}) = L_1(m_1) + L_5(m_5) + L_4(m_4) = L_1(m_1) + L_4(m_4)$, *and* $L^3(\boldsymbol{m}) = L_3(m_3) + L_4(m_4)$.

The Wardrop (Nash) equilibrium for the Braess network in Example 7 is that all the n players travel along path 2 since it is the weakly dominant path for any player as the expected latency incurred along path 2 is always less than or equal to the expected latency incurred along the other two paths 1 and 3,

$$\begin{aligned} l^2(\boldsymbol{m}) &= l_1(m_1) + l_5(m_5) + l_4(m_4) \\ &= \frac{m_1}{n} + \frac{m_4}{n} \begin{cases} \leq \frac{m_1}{n} + 1 = l_1(m_1) + l_2(m_2) = l^1(\boldsymbol{m}), \\ \leq 1 + \frac{m_4}{n} = l_3(m_3) + l_4(m_4) = l^3(\boldsymbol{m}). \end{cases} \end{aligned}$$

As a result, the Wardrop equilibrium for Braess network is $\boldsymbol{p}^*_W = (2, 2, \ldots, 2)$ with social delay $D_W(\boldsymbol{p}^*_W) = 2$. However, although path 2 has latency less than or equal to that of paths 1 and 3, $l^2(\boldsymbol{m}) \leq \big(l^1(\boldsymbol{m}), l^3(\boldsymbol{m})\big)$, the variance of travel time along path 2 at full capacity is larger than that along paths 1 and 3, which increases the risk and uncertainty of traveling along path 2. In fact, path 1 (or 3) with higher expected travel time is more likely to have a lower delay than the rest of the paths; i.e., $P\Big(L^1(0) \leq \big(L^2(n), L^3(0)\big)\Big) = \frac{3}{8} > \frac{1}{4} = P\Big(L^2(n) \leq \big(L^1(0), L^3(0)\big)\Big)$. As a result, a risk-averse player selects paths 1 or 3 for commute when path 2 is at full capacity, and as is shown later, the risk-averse behavior of players decreases social delay for this example.

4.2.2 R-ABADI Equilibrium

In the classical Wardrop (Nash) equilibrium, the best response of player $i \in [n]$ to the mixed strategy $\boldsymbol{\sigma}_{-i}$ of the other $[n] \setminus i$ players is defined as the set

$$\arg\min_{p_i \in \mathcal{P}_i} \bar{l}^i(p_i, \boldsymbol{\sigma}_{-i}).$$

In other words, the best response for player i given $\boldsymbol{\sigma}_{-i}$ is defined as the path that minimizes the expected travel time. However, motivated by Examples 6 and 7, the path with minimum expected latency may have a high volatility as well that causes risky scenarios for travelers. As a result, the classical Wardrop (Nash) equilibrium that ignores the distribution of path latency except for taking the expected latency into account, that does not carry any information about variance and the shape of the distribution, falls short in addressing risk-averse behavior of players. In this chapter, motivated by Examples 6 and 7, we propose a Risk-Averse Best Action Decision with Incomplete Information (R-ABADI) of a player to the strategy of the other players in a stochastic congestion game as follows. Note that the risk-averse best-response/equilibrium and R-ABADI best-response/equilibrium are used interchangeably throughout this chapter.

Definition 11. *Given the mixed strategy profile $\boldsymbol{\sigma}_{-i}$ of players $[n] \setminus i$, the set of mixed strategy risk-averse R-ABADI best responses of player i is the set of all probability distributions over the set*

$$\underset{p_i \in \mathcal{P}_i}{\arg\max}\ P\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p_i, \boldsymbol{\sigma}_{-i})\right), \tag{4.3}$$

where what we mean by $\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})$ being less than or equal to $\overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p_i, \boldsymbol{\sigma}_{-i})$ when $\mathcal{P}_i \setminus p_i \neq \varnothing$ is that $\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})$ is less than or equal to $\overline{L}^i(p_i', \boldsymbol{\sigma}_{-i})$ for all $p_i' \in \mathcal{P}_i \setminus p_i$; otherwise, if $\mathcal{P}_i \setminus p_i = \varnothing$, player i only has a single option that can be played. The same randomness on the action of players $[n] \setminus i$ is considered in $\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})$ for all $p_i \in \mathcal{P}_i$. Given the mixed strategy $\boldsymbol{\sigma}_{-i}$ of players $[n] \setminus i$, the risk-averse best response set of player i's strategies is denoted by $RB(\boldsymbol{\sigma}_{-i})$, which is in general a set-valued function.

The risk-averse equilibrium for stochastic congestion games is defined as follows.

Definition 12. *A strategy profile $\boldsymbol{\sigma}^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_N^*)$ is a risk-averse R-ABADI equilibrium if and only if $\sigma_i^* \in RB(\boldsymbol{\sigma}_{-i}^*)$ for all $i \in [n]$.*

The following theorem, which is a special case of Theorem 4, proves the existence of a risk-averse equilibrium for any stochastic congestion game with finite number of players and pure strategy sets $\mathcal{P}_i$ for all $i \in [n]$ with finite cardinality.

Theorem 5. *For any finite n-player stochastic congestion game, a risk-averse equilibrium exists.*

The proof of Theorem 5 is provided in Appendix C.1.

As a direct result of Definitions 11 and 12, the pure strategy risk-averse best response and pure strategy risk-averse equilibrium are defined as follows. The pure strategy risk-averse best response of player i to the pure strategy $\boldsymbol{p}_{-i}$ of players $[n] \setminus i$ is the set

$$\begin{cases} \arg\max_{p_i \in \mathcal{P}_i} P\Big(L^i\left(p_i, \boldsymbol{p}_{-i}\right) \leq \boldsymbol{L}^i\left(\mathcal{P}_i \setminus p_i, \boldsymbol{p}_{-i}\right)\Big), & \text{if } \mathcal{P}_i \setminus p_i \neq \varnothing, \\ p_i, \quad \text{if } \mathcal{P}_i \setminus p_i = \varnothing. \end{cases} \tag{4.4}$$

Given the pure strategy $\boldsymbol{p}_{-i}$ of players $[n] \setminus i$, the risk-averse best response set of player i in Equation (4.4) is denoted by $RB(\boldsymbol{p}_{-i})$ (overloading notation, $RB(.)$ is used for both pure and mixed strategy risk-averse best responses). As a result, a pure strategy profile $\boldsymbol{p}^* = (p_1^*, p_2^*, \ldots, p_n^*)$ is a pure strategy risk-averse equilibrium if and only if $p_i^* \in RB(\boldsymbol{p}_{-i}^*)$ for all $i \in [n]$.

Strict dominance in the classical Wardrop (Nash) equilibrium is defined as follows. A pure strategy $p_i \in \mathcal{P}_i$ of player i strictly dominates a second pure strategy $p_i' \in \mathcal{P}_i$ of the player if

$$l^i(p_i, \boldsymbol{p}_{-i}) < l^i(p_i', \boldsymbol{p}_{-i}),\ \forall \boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}.$$

The solution concept of iterated elimination of strictly dominated strategies can also be applied to the risk-averse equilibrium using the following definition.

Definition 13. *A pure strategy $p_i \in \mathcal{P}_i$ of player i strictly dominates a second pure strategy $p_i' \in \mathcal{P}_i$ of the player in the risk-averse equilibrium if*

$$\begin{aligned} &P\Big(L^i\left(p_i, \boldsymbol{p}_{-i}\right) \leq \boldsymbol{L}^i\left(\mathcal{P}_i \setminus p_i, \boldsymbol{p}_{-i}\right)\Big) \\ >&P\Big(L^i\left(p_i', \boldsymbol{p}_{-i}\right) \leq \boldsymbol{L}^i\left(\mathcal{P}_i \setminus p_i', \boldsymbol{p}_{-i}\right)\Big),\ \forall \boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}. \end{aligned} \tag{4.5}$$

Consider path $p_i \in \mathcal{P}_i$ strictly dominates path $p_i' \in \mathcal{P}_i$ for player i; then,

for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$

$$
\begin{aligned}
& P\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p_i, \boldsymbol{\sigma}_{-i})\right) \\
& \overset{(a)}{=} \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(P\left(L^i(p_i, \boldsymbol{p}_{-i}) \leq \boldsymbol{L}^i(\mathcal{P}_i \setminus p_i, \boldsymbol{p}_{-i})\right) \cdot \boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \right) \\
& \overset{(b)}{>} \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(P\left(L^i(p'_i, \boldsymbol{p}_{-i}) \leq \boldsymbol{L}^i(\mathcal{P}_i \setminus p'_i, \boldsymbol{p}_{-i})\right) \cdot \boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \right) \\
& = P\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p'_i, \boldsymbol{\sigma}_{-i})\right),
\end{aligned}
\tag{4.6}
$$

where (a) is true by the law of total probability, $\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) = \prod_{j \in [n] \setminus i} \sigma_j(p_j)$ and p_j is the corresponding strategy of player j in $\boldsymbol{p}_{-i}$, and (b) is followed by Equation (4.5) in Definition 13. By Equation (4.6) and Equation (4.3) in Definition 11, a strictly dominated pure strategy cannot be a best response to any mixed strategy profile $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$, so it can be removed from the set of strategies of player i.

In order to find the risk-averse equilibrium for a stochastic congestion game, we use support enumeration. For example, hypothesize that $\boldsymbol{\mathcal{P}'} := \{\mathcal{P}'_1, \mathcal{P}'_2, \ldots, \mathcal{P}'_n\}$ is the *support* of a risk-averse equilibrium, where $\mathcal{P}'_i$ is the set of pure strategies of player i that are played with non-zero probability and $\sigma_i(p_i)$ for $p_i \in \mathcal{P}'_i$ indicates the probability mass function on the support. At equilibrium, player $i \in [n]$ should be indifferent between strategies in the set $\mathcal{P}'_i$, has no incentive to deviate to the rest of strategies in the set $\mathcal{P}_i \setminus \mathcal{P}'_i$, and the probability mass function over the support should add to one. As a result, if there is a risk-averse equilibrium with the mentioned support, it is the solution of the following set of equations for $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$:

$$
\begin{cases}
P\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p'_i, \boldsymbol{\sigma}_{-i})\right) \\
\geq P\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p_i, \boldsymbol{\sigma}_{-i})\right), \forall p_i \in \mathcal{P}_i, p'_i \in \mathcal{P}'_i, \forall i \in [n], \\
\\
\sum_{p_i \in \mathcal{P}'_i} \sigma_i(p_i) = 1, \forall i \in [n], \\
\\
\sigma_i(p_i) = 0, \forall p_i \in \mathcal{P}_i \setminus \mathcal{P}'_i, \forall i \in [n].
\end{cases}
\tag{4.7}
$$

As mentioned earlier in Equation (4.6), using the law of total probability,

we have

$$\begin{aligned} & P\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p_i, \boldsymbol{\sigma}_{-i})\right) \\ = & \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(P\left(L^i(p_i, \boldsymbol{p}_{-i}) \leq \boldsymbol{L}^i(\mathcal{P}_i \setminus p_i, \boldsymbol{p}_{-i})\right) \cdot \boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \right) \\ = & \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \Big(t_i(p_i, \boldsymbol{p}_{-i}) \cdot \boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \Big), \end{aligned} \tag{4.8}$$

where $t_i(p_i, \boldsymbol{p}_{-i}) \coloneqq P\left(L^i(p_i, \boldsymbol{p}_{-i}) \leq \boldsymbol{L}^i(\mathcal{P}_i \setminus p_i, \boldsymbol{p}_{-i})\right)$ is the i-th element of an n-dimensional vector called $\boldsymbol{t}(p_i, \boldsymbol{p}_{-i})$. Construct a *risk-averse probability tensor* of rank n where $\mathcal{P}_i$ forms the i-th dimension of the tensor. Let the element associated with $(p_i, \boldsymbol{p}_{-i})$ in the tensor be the vector $\boldsymbol{t}(p_i, \boldsymbol{p}_{-i})$. Equations (4.7) and (4.8) along with the definition of the risk-averse probability tensor provide us with an alternative approach for deriving the risk-averse equilibrium, which is to find the Wardrop (Nash) equilibrium on the risk-averse probability tensor.

The mean-variance (MV) and conditional value at risk level α (CVaR_α) methods are two well-known frameworks to consider risk in statistics. In the next two subsections, two new risk-averse equilibria based on these two concepts are proposed.

4.2.3 Mean-Variance Equilibrium

As seen in Examples 6 and 7, the high variance of paths with lower expected travel time can result in uncertainty and impose high latency for travelers. The mean-variance framework in statistics addresses this issue by keeping a balance between low latency and low variance. Applying this method to the proposed stochastic congestion game setting, the mean-variance best response and mean-variance equilibrium are defined as follows.

Definition 14. *Given the mixed strategy profile $\boldsymbol{\sigma}_{-i}$ of players $[n] \setminus i$, the set of mixed strategy mean-variance best responses of player i is the set of all probability distributions over the set*

$$\underset{p_i \in \mathcal{P}_i}{\arg\min}\ \text{Var}\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p_i, \boldsymbol{\sigma}_{-i}), \tag{4.9}$$

where the variance $\text{Var}\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right)$ *can be calculated using the pdf of the random variable* $\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})$ *provided in Equation* (4.1) *and* $\rho \geq 0$ *is a hyper-parameter capturing the absolute risk tolerance. Given the mixed strategy* $\boldsymbol{\sigma}_{-i}$ *of players* $[n] \setminus i$*, the mean-variance best response set of player* i*'s strategies is denoted by* $MB(\boldsymbol{\sigma}_{-i})$*, which is in general a set-valued function.*

Definition 15. *A strategy profile* $\boldsymbol{\sigma}^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_N^*)$ *is a mean-variance equilibrium if and only if* $\sigma_i^* \in MB(\boldsymbol{\sigma}_{-i}^*)$ *for all* $i \in [n]$*.*

The existence of the mean-variance equilibrium is discussed in the following theorem.

Theorem 6. *For any finite n-player stochastic congestion game, a mean-variance equilibrium exists.*

The proof of Theorem 6 is provided in Appendix C.2.

The pure strategy mean-variance best response of player i to the pure strategy $\boldsymbol{p}_{-i}$ of players $[n] \setminus i$ is the set

$$\underset{p_i \in \mathcal{P}_i}{\arg\min}\ \text{Var}\left(L^i(p_i, \boldsymbol{p}_{-i})\right) + \rho \cdot l^i(p_i, \boldsymbol{p}_{-i}), \tag{4.10}$$

where

$$\begin{aligned}\text{Var}\left(L^i(p_i, \boldsymbol{p}_{-i})\right) &= \text{Var}\left(\sum_{e \in p_i} L_e(p_i, \boldsymbol{p}_{-i})\right) \\ &= \sum_{e \in p_i} \sum_{e' \in p_i} \text{Cov}\left(L_e(p_i, \boldsymbol{p}_{-i}), L_{e'}(p_i, \boldsymbol{p}_{-i})\right).\end{aligned}$$

Given the pure strategy $\boldsymbol{p}_{-i}$ of players $[n] \setminus i$, the mean-variance best response set of player i in Equation (4.10) is denoted by $MB(\boldsymbol{p}_{-i})$ (overloading notation, $MB(.)$ is used for both pure and mixed strategy mean-variance best responses). As a result, a pure strategy profile $\boldsymbol{p}^* = (p_1^*, p_2^*, \ldots, p_n^*)$ is a pure strategy mean-variance equilibrium if and only if $p_i^* \in MB(\boldsymbol{p}_{-i}^*)$ for all $i \in [n]$. The strict dominance concept is straightforward among pure strategy profiles in mean-variance equilibrium that is defined as follows. A pure strategy $p_i \in \mathcal{P}_i$ of player i strictly dominates a second pure strategy $p_i' \in \mathcal{P}_i$ of the player in pure strategy mean-variance equilibrium if

$$\begin{aligned}&\text{Var}\left(L^i(p_i, \boldsymbol{p}_{-i})\right) + \rho \cdot l^i(p_i, \boldsymbol{p}_{-i}) \\ &< \text{Var}\left(L^i(p_i', \boldsymbol{p}_{-i})\right) + \rho \cdot l^i(p_i', \boldsymbol{p}_{-i}),\ \forall \boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}.\end{aligned} \tag{4.11}$$

However, due to the fact that variance is not a linear operator, strict dominance may not be derived from Equation (4.11) for mixed strategy mean-variance equilibrium as described below.

$$
\begin{aligned}
&\mathrm{Var}\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p_i, \boldsymbol{\sigma}_{-i}) \\
&\overset{(a)}{=} \mathrm{E}\left[\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right)^2\right] - \left(\overline{l}^i(p_i, \boldsymbol{\sigma}_{-i})\right)^2 + \rho \cdot \overline{l}^i(p_i, \boldsymbol{\sigma}_{-i}) \\
&\overset{(b)}{=} \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot \mathrm{E}\left[\left(L^i(p_i, \boldsymbol{p}_{-i})\right)^2\right]\right) \\
&\quad - \left(\sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot l^i(p_i, \boldsymbol{p}_{-i})\right)\right)^2 + \rho \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot l^i(p_i, \boldsymbol{p}_{-i})\right) \\
&\overset{(c)}{=} \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot \mathrm{E}\left[\left(L^i(p_i, \boldsymbol{p}_{-i})\right)^2\right]\right) \\
&\quad - \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\sum_{\boldsymbol{p}'_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot \boldsymbol{\sigma}(\boldsymbol{p}'_{-i}) \cdot l^i(p_i, \boldsymbol{p}_{-i}) \cdot l^i(p_i, \boldsymbol{p}'_{-i})\right)\right) \\
&\quad + \rho \cdot \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot l^i(p_i, \boldsymbol{p}_{-i})\right) \\
&\overset{(d)}{=} \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot \left(\mathrm{E}\left[\left(L^i(p_i, \boldsymbol{p}_{-i})\right)^2\right] - l^i(p_i, \boldsymbol{p}_{-i}) \times \right. \\
&\qquad \left. \sum_{\boldsymbol{p}'_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}'_{-i}) \cdot l^i(p_i, \boldsymbol{p}'_{-i})\right) + \rho \cdot l^i(p_i, \boldsymbol{p}_{-i})\right) \\
&= \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot \left(\mathrm{E}\left[\left(L^i(p_i, \boldsymbol{p}_{-i})\right)^2\right] - l^i(p_i, \boldsymbol{p}_{-i}) \times \right. \\
&\qquad \left. \left(\sum_{\boldsymbol{p}'_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}'_{-i}) \cdot l^i(p_i, \boldsymbol{p}'_{-i})\right) + \rho\right)\right),
\end{aligned}
\tag{4.12}
$$

where (a) is true by the definition of variance, (b) is followed by Equation (4.2), (c) is derived by expanding the second term, and (d) is true by combining the summation over $\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}$ and factoring $\boldsymbol{\sigma}(\boldsymbol{p}_{-i})$. As can be seen in Equation (4.12), since variance is a non-linear operator, it is not clear whether Equation (4.11) can result in $\mathrm{Var}\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p_i, \boldsymbol{\sigma}_{-i}) < \mathrm{Var}\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p'_i, \boldsymbol{\sigma}_{-i})$ for all $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$. As a result, use of strict

dominance in the mixed strategy mean-variance equilibrium is not advised. In certain circumstances though, we can propose conditions for strict dominance; e.g., when $l^i(\boldsymbol{p}) \leq \frac{\rho}{2}$ for all $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$ and for all $i \in [n]$ which is discussed in the following definition or when $l^i(\boldsymbol{p}) \geq \frac{\rho}{2}$ for all $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$ and for all $i \in [n]$.

Definition 16. *Suppose $l^i(\boldsymbol{p}) \leq \frac{\rho}{2}$ for all $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$ and for all $i \in [n]$. Then, pure strategy $p_i \in \mathcal{P}_i$ of player i strictly dominates a second pure strategy $p_i' \in \mathcal{P}_i$ of the player in the mean-variance equilibrium if*

$$l^i\left(p_i, \boldsymbol{p}_{-i}\right) < l^i\left(p_i', \boldsymbol{p}_{-i}\right), \ \forall \boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}, \tag{4.13}$$

and

$$\mathrm{E}\left[\left(L^i\left(p_i, \boldsymbol{p}_{-i}\right)\right)^2\right] < \mathrm{E}\left[\left(L^i\left(p_i', \boldsymbol{p}_{-i}\right)\right)^2\right], \ \forall \boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}. \tag{4.14}$$

Consider that path $p_i \in \mathcal{P}_i$ strictly dominates path $p_i' \in \mathcal{P}_i$ for player i as defined in Definition 16; then, using Equation (4.13), for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$,

$$\begin{aligned} \overline{l}^i(p_i, \boldsymbol{\sigma}_{-i}) &= \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot l^i(p_i, \boldsymbol{p}_{-i})\right) \\ &< \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot l^i(p_i', \boldsymbol{p}_{-i})\right) = \overline{l}^i(p_i', \boldsymbol{\sigma}_{-i}). \end{aligned} \tag{4.15}$$

Note that $\overline{l}^i(p_i, \boldsymbol{\sigma}_{-i}) \leq \frac{\rho}{2}$ for all $p_i \in \mathcal{P}_i$, for all $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$, and for all $i \in [n]$ as a result of $l^i(\boldsymbol{p}) \leq \frac{\rho}{2}$ for all $\boldsymbol{p} \in \boldsymbol{\mathcal{P}}$ and for all $i \in [n]$. Hence, using the fact that the function $-f^2 + \rho \cdot f$ is increasing for $f \leq \frac{\rho}{2}$, for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$ we have

$$\begin{aligned} &-\left(\overline{l}^i(p_i, \boldsymbol{\sigma}_{-i})\right)^2 + \rho \cdot \overline{l}^i(p_i, \boldsymbol{\sigma}_{-i}) \\ &< -\left(\overline{l}^i(p_i', \boldsymbol{\sigma}_{-i})\right)^2 + \rho \cdot \overline{l}^i(p_i', \boldsymbol{\sigma}_{-i}). \end{aligned} \tag{4.16}$$

On the other hand, using Equation (4.14), we have

$$\begin{aligned} \mathrm{E}\left[\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right)^2\right] &= \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot \mathrm{E}\left[\left(L^i(p_i, \boldsymbol{p}_{-i})\right)^2\right]\right) \\ < \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot \mathrm{E}\left[\left(L^i(p_i', \boldsymbol{p}_{-i})\right)^2\right]\right) &= \mathrm{E}\left[\left(\overline{L}^i(p_i', \boldsymbol{\sigma}_{-i})\right)^2\right]. \end{aligned} \tag{4.17}$$

Finally, Equations (4.16) and (4.17) conclude that $\mathrm{Var}\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot$

$\overline{l}^i(p_i, \boldsymbol{\sigma}_{-i}) < \text{Var}\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p'_i, \boldsymbol{\sigma}_{-i})$ for all $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$.

In order to find the mean-variance equilibrium for a stochastic congestion game, we use support enumeration. For example, hypothesize $\boldsymbol{\mathcal{P}'} := \{\mathcal{P}'_1, \mathcal{P}'_2, \ldots, \mathcal{P}'_n\}$ to be the support of a mean-variance equilibrium, where $\mathcal{P}'_i$ is the set of pure strategies of player i that are played with non-zero probability and $\sigma_i(p_i)$ for $p_i \in \mathcal{P}'_i$ indicates the probability mass function on the support. At equilibrium, player $i \in [n]$ should be indifferent between strategies in the set $\mathcal{P}'_i$, has no incentive to deviate to the rest of strategies in the set $\mathcal{P}_i \setminus \mathcal{P}'_i$, and the probability mass function over the support should add to one. As a result, if there is a mean-variance equilibrium with the mentioned support, it is the solution of the following set of equations for $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$:

$$\begin{cases} \text{Var}\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p'_i, \boldsymbol{\sigma}_{-i}) \\ \leq \text{Var}\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p_i, \boldsymbol{\sigma}_{-i}), \forall p_i \in \mathcal{P}_i, p'_i \in \mathcal{P}'_i, \forall i \in [n], \\ \\ \sum_{p_i \in \mathcal{P}'_i} \sigma_i(p_i) = 1, \forall i \in [n], \\ \\ \sigma_i(p_i) = 0, \forall p_i \in \mathcal{P}_i \setminus \mathcal{P}'_i, \forall i \in [n]. \end{cases} \tag{4.18}$$

4.2.4 CVaR$_\alpha$ Equilibrium

The conditional value at risk level α (CVaR$_\alpha$) is another framework in statistics to measure risk and to address the risk-averse behavior. Applying this method to the proposed stochastic congestion game setting, the CVaR$_\alpha$ best response and CVaR$_\alpha$ equilibrium are defined below.

Definition 17. *Given the mixed strategy profile $\boldsymbol{\sigma}_{-i}$ of players $[n] \setminus i$, the set of mixed strategy CVaR$_\alpha$ best responses of player i is the set of all probability distributions over the set*

$$\begin{aligned} &\arg\min_{p_i \in \mathcal{P}_i} \; CVaR_\alpha\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right) \\ = &\arg\min_{p_i \in \mathcal{P}_i} \; \text{E}\left[\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \geq v^i_\alpha(p_i, \boldsymbol{\sigma}_{-i})\right], \end{aligned} \tag{4.19}$$

where $v^i_\alpha(p_i, \boldsymbol{\sigma}_{-i})$ is a constant derived by solving the equality $P\big(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \geq v^i_\alpha(p_i, \boldsymbol{\sigma}_{-i})\big) = \alpha$ and the constant $0 < \alpha \leq 1$ is a hyper-parameter depicting

the risk level. Given the mixed strategy $\boldsymbol{\sigma}_{-i}$ of players $[n] \setminus i$, the $CVaR_\alpha$ best response set of player i's strategies is denoted by $CB(\boldsymbol{\sigma}_{-i})$, which is in general a set-valued function.

Definition 18. *A strategy profile $\boldsymbol{\sigma}^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_N^*)$ is a $CVaR_\alpha$ equilibrium if and only if $\sigma_i^* \in CB(\boldsymbol{\sigma}_{-i}^*)$ for all $i \in [n]$.*

The existence of the CVaR_α equilibrium is discussed in the following theorem.

Theorem 7. *For any finite n-player stochastic congestion game, a $CVaR_\alpha$ equilibrium exists.*

The proof of Theorem 7 is provided in Appendix C.3.

The pure strategy CVaR_α best response of player i to the pure strategy $\boldsymbol{p}_{-i}$ of players $[n] \setminus i$ is the set

$$\begin{aligned} &\underset{p_i \in \mathcal{P}_i}{\arg\min}\ CVaR_\alpha\left(L^i(p_i, \boldsymbol{p}_{-i})\right) \\ &= \underset{p_i \in \mathcal{P}_i}{\arg\min}\ \mathrm{E}\left[L^i(p_i, \boldsymbol{p}_{-i}) \middle| L^i(p_i, \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{p}_{-i})\right], \end{aligned} \tag{4.20}$$

where $v_\alpha^i(p_i, \boldsymbol{p}_{-i})$ is a constant derived by solving the equality $P\big(L^i(p_i, \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{p}_{-i})\big) = \alpha$ and the constant $0 < \alpha \leq 1$ is the hyper-parameter depicting risk level. Given the pure strategy $\boldsymbol{p}_{-i}$ of players $[n] \setminus i$, the CVaR_α best response set of player i in Equation (4.20) is denoted by $CB(\boldsymbol{p}_{-i})$ (overloading notation, $CB(.)$ is used for both pure and mixed strategy CVaR_α best responses). As a result, a pure strategy profile $\boldsymbol{p}^* = (p_1^*, p_2^*, \ldots, p_n^*)$ is a pure strategy CVaR_α equilibrium if and only if $p_i^* \in CB(\boldsymbol{p}_{-i}^*)$ for all $i \in [n]$. A pure strategy $p_i \in \mathcal{P}_i$ of player i strictly dominates a second pure strategy $p_i' \in \mathcal{P}_i$ of the player in pure strategy CVaR_α equilibrium if

$$\begin{aligned} &\mathrm{E}\left[L^i(p_i, \boldsymbol{p}_{-i}) \middle| L^i(p_i, \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{p}_{-i})\right] \\ &< \mathrm{E}\left[L^i(p_i', \boldsymbol{p}_{-i}) \middle| L^i(p_i', \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i', \boldsymbol{p}_{-i})\right],\ \forall \boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}, \end{aligned} \tag{4.21}$$

where $v_\alpha^i(p_i, \boldsymbol{p}_{-i})$ and $v_\alpha^i(p_i', \boldsymbol{p}_{-i})$ are constants derived by solving the equation $P\big(L^i(p_i, \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{p}_{-i})\big) = \alpha$ and $P\left(L^i(p_i', \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i', \boldsymbol{p}_{-i})\right) = \alpha$, and the constant $0 < \alpha \leq 1$ is the risk level hyper-parameter. However, similar to the mean-variance equilibrium, strict dominance may not be derived

from Equation (4.21) for mixed strategy CVaR_α equilibrium as described below. Using Equation (4.1) that provides the pdf function of the random variable and $P\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\right) = \alpha$, the distribution of the random variable $\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\right)$ is

$$\left(\sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(f^i(x|(p_i, \boldsymbol{p}_{-i})) \cdot \boldsymbol{\sigma}(\boldsymbol{p}_{-i})\right) \Big/ \alpha\right) \cdot \mathbb{1}\left\{x \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\right\}. \tag{4.22}$$

As a result,

$$\begin{aligned}
&\text{E}\left[\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\right] \\
&\overset{(a)}{=} \frac{1}{\alpha} \cdot \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \left(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot \int_{-\infty}^{\infty} \left(x \cdot f^i(x|(p_i, \boldsymbol{p}_{-i})) \cdot \mathbb{1}\left\{x \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\right\}\right) dx\right) \\
&\overset{(b)}{=} \frac{1}{\alpha} \cdot \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \Bigg(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot P\big(L^i(p_i, \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\big) \times \\
&\qquad\qquad \int_{v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})}^{\infty} \left(x \cdot \frac{f^i(x|(p_i, \boldsymbol{p}_{-i}))}{P\big(L^i(p_i, \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\big)}\right) dx\Bigg) \\
&= \frac{1}{\alpha} \cdot \sum_{\boldsymbol{p}_{-i} \in \boldsymbol{\mathcal{P}}_{-i}} \Bigg(\boldsymbol{\sigma}(\boldsymbol{p}_{-i}) \cdot P\big(L^i(p_i, \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\big) \times \\
&\qquad\qquad \text{E}\left[L^i(p_i, \boldsymbol{p}_{-i}) \middle| L^i(p_i, \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\right]\Bigg),
\end{aligned} \tag{4.23}$$

where (a) is true by using the pdf of the corresponding random variable in Equation (4.22) and switching the order of summation and integral and (b) is true by multiplying and dividing by the term $P\big(L^i(p_i, \boldsymbol{p}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\big)$. As can be seen in Equation (4.23), it is not clear whether Equation (4.21) can result in the equation of interest $\text{E}\left[\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\right] < \text{E}\left[\overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(p_i', \boldsymbol{\sigma}_{-i})\right]$ for all $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$. As a result, use of strict dominance in the mixed strategy CVaR_α equilibrium is not advised due to its complication.

In order to find the CVaR_α equilibrium for a stochastic congestion game, we use support enumeration. For example, hypothesize $\boldsymbol{\mathcal{P}}' \coloneqq \{\mathcal{P}_1', \mathcal{P}_2', \ldots, \mathcal{P}_n'\}$ to be the support of a CVaR_α equilibrium, where $\mathcal{P}_i'$ is the set of pure strategies of player i that are played with non-zero probability and $\sigma_i(p_i)$

for $p_i \in \mathcal{P}'_i$ indicates the probability mass function on the support. At equilibrium, player $i \in [n]$ should be indifferent between strategies in the set $\mathcal{P}'_i$, has no incentive to deviate to the rest of strategies in the set $\mathcal{P}_i \setminus \mathcal{P}'_i$, and the probability mass function over the support should add to one. As a result, if there is a CVaR_α equilibrium with the mentioned support, it is the solution of the following set of equations for $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$:

$$\begin{cases} \mathrm{E}\left[\overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i}) \geq v^i_\alpha(p'_i, \boldsymbol{\sigma}_{-i})\right] \\ \leq \mathrm{E}\left[\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \geq v^i_\alpha(p_i, \boldsymbol{\sigma}_{-i})\right], \forall p_i \in \mathcal{P}_i, p'_i \in \mathcal{P}'_i, \forall i \in [n], \\ \\ \sum_{p_i \in \mathcal{P}'_i} \sigma_i(p_i) = 1, \forall i \in [n], \\ \\ \sigma_i(p_i) = 0, \forall p_i \in \mathcal{P}_i \setminus \mathcal{P}'_i, \forall i \in [n]. \end{cases} \tag{4.24}$$

Remark 7. *It is noteworthy that the polynomial terms in Equation* (4.7) *for the risk-averse equilibrium are of degree* $n-1$ *while the polynomial terms in Equation* (4.18) *for the mean-variance equilibrium are of degree* $2(n-1)$ *for* n *number of players. On the other hand, it is more complicated to solve for Equation* (4.24) *as the top* α *quantile of distributions should be calculated.*

4.3 Numerical Results

The risk-averse, mean-variance, and CVaR_α equilibria are numerically analyzed for Examples 6 and 7 in this section. The price of anarchy for each of the mentioned equilibria is calculated as well. In the end, extra examples are presented to shed light on the corner cases of each one of the equilibria and to provide insight on how to tackle such circumstances.

In order to find any of the three types of pure equilibria for the Pigou network in Example 6 with n players, hypothesize that m_1 players choose link 1 and $m_2 = n - m_1$ players choose link 2 and check whether any players has any incentive in the corresponding sense of the equilibrium of the interest to change route, given the pure strategy of the other players. If none of the players has any incentive to change route given the pure strategy of the rest of players, $(m_1, n - m_1)$ is a pure equilibrium, where (m_1, m_2) denotes that m_1 players select link 1 and m_2 players select link 2. By varying m_1 from zero

to n and taking the above procedure, the pure equilibrium is found if any exists. Given a fixed number of players m_1 that choose link 1, it is obvious that they all have the same incentive to change to link 2 or stay in link 1, and all of the $m_2 = n - m_1$ players have the same incentive to change to link 1 or stay in link 2. As a result, if a specific player out of the m_1 players has no incentive to switch to link 2 given the pure strategy of the other players, and a specific player out of the m_2 players has no incentive to switch to link 1 given the pure strategy of the other players, $(m_1, m_2 = n - m_1)$ is a pure equilibrium. In other words, $(m_1, m_2 = n - m_1)$ is a pure risk-averse equilibrium if

$$\begin{cases} P\big(L_1(m_1) \leq L_2(m_2 + 1)\big) \geq 0.5, \\ \\ P\big(L_2(m_2) \leq L_1(m_1 + 1)\big) \geq 0.5, \end{cases} \tag{4.25}$$

where the first inequality is true since each player has two options, link 1 and link 2, so $P\big(L_1(m_1) \leq L_2(m_2 + 1)\big) \geq P\big(L_2(m_2 + 1) \leq L_1(m_1)\big)$, and since random variables are continuous we have $P\big(L_1(m_1) \leq L_2(m_2 + 1)\big) + P\big(L_2(m_2 + 1) \leq L_1(m_1)\big) = 1$, which results in $P\big(L_1(m_1) \leq L_2(m_2 + 1)\big) \geq 0.5$. The second inequality is true due to a similar reasoning. By varying m_1 from zero to n, if Equation (4.25) holds for $(m_1, m_2 = n - m_1)$, it is a pure risk-averse equilibrium.

Similar to the above approach, $(m_1, m_2 = n - m_1)$ is a pure mean-variance equilibrium if

$$\begin{cases} \mathrm{Var}\big(L_1(m_1)\big) + \rho \cdot l_1(m_1) \leq \mathrm{Var}\big(L_2(m_2 + 1)\big) + \rho \cdot l_2(m_2 + 1), \\ \\ \mathrm{Var}\big(L_2(m_2)\big) + \rho \cdot l_2(m_2) \leq \mathrm{Var}\big(L_1(m_1 + 1)\big) + \rho \cdot l_1(m_1 + 1). \end{cases} \tag{4.26}$$

Again, by varying m_1 from zero to n, if Equation (4.26) holds for $(m_1, m_2 = n - m_1)$, it is a pure mean-variance equilibrium. Similarly, $(m_1, m_2 = n - m_1)$ is a pure CVaR_α equilibrium if

$$\begin{cases} \mathrm{E}\big[L_1(m_1) \big| L_1(m_1) \geq v_\alpha^1(m_1)\big] \leq \mathrm{E}\big[L_2(m_2 + 1) \big| L_2(m_2 + 1) \geq v_\alpha^2(m_2 + 1)\big], \\ \\ \mathrm{E}\big[L_2(m_2) \big| L_2(m_2) \geq v_\alpha^2(m_2)\big] \leq \mathrm{E}\big[L_1(m_1 + 1) \big| L_1(m_1 + 1) \geq v_\alpha^1(m_1 + 1)\big], \end{cases} \tag{4.27}$$

where $P\big(L_1(m_1) \geq v_\alpha^1(m_1)\big) = P\big(L_2(m_2 + 1) \geq v_\alpha^2(m_2 + 1)\big) = P\big(L_2(m_2) \geq$

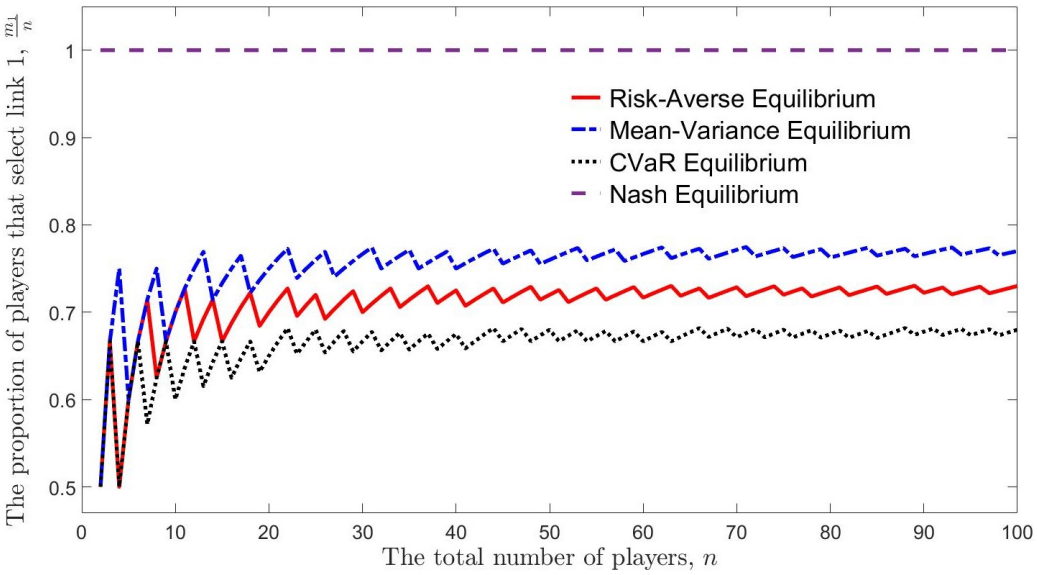

Figure 4.3: The pure risk-averse, mean-variance ($\rho = 1$), CVaR_α ($\alpha = 0.1$), and Nash equilibria of the Pigou network in Example 6 are denoted for different numbers of players.

$v_\alpha^2(m_2)\big) = P\big(L_1(m_1+1) \geq v_\alpha^1(m_1+1)\big) = \alpha$. By varying m_1 from zero to n, if Equation (4.27) holds for $(m_1, m_2 = n - m_1)$, it is a pure CVaR_α equilibrium.

Note that the equilibrium in the Pigou network in Example 6 is characterized by m_1, since m_2 can be derived given m_1. The pure risk-averse, mean-variance ($\rho = 1$), and CVaR_α ($\alpha = 0.1$) equilibria are found for the mentioned Pigou network and the proportion of players who select link 1; i.e., $\frac{m_1}{n}$, is depicted in Figure 4.3 for different values of n. Under the Nash equilibrium, no matter what the probability distributions of latency over links look like, all players select link 1 as it has less or equal latency in expectation. Hence, $(n, 0)$ is the Nash equilibrium for all n, which corresponds to $\frac{m_1}{n} = 1$ as depicted in Figure 4.3.

The social delay/latency defined as the expected average delay/latency incurred by the n players in the Pigou network in Example 6 under the pure strategy (m_1, m_2) is $D(m_1) = \frac{1}{n}\left(m_1 \cdot \frac{m_1}{n} + (n - m_1)\right) = \left(\frac{m_1}{n}\right)^2 - \frac{m_1}{n} + 1$, which is minimized when $m_1 = \frac{n}{2}$ for an even n, and $m_1 = \lfloor \frac{n}{2} \rfloor$ and $m_1 = \lceil \frac{n}{2} \rceil$ for an odd n. As a result, it is socially optimal that about half of the players take the top link and the rest take the bottom link to travel from source to destination in the Pigou network, which results in a social latency

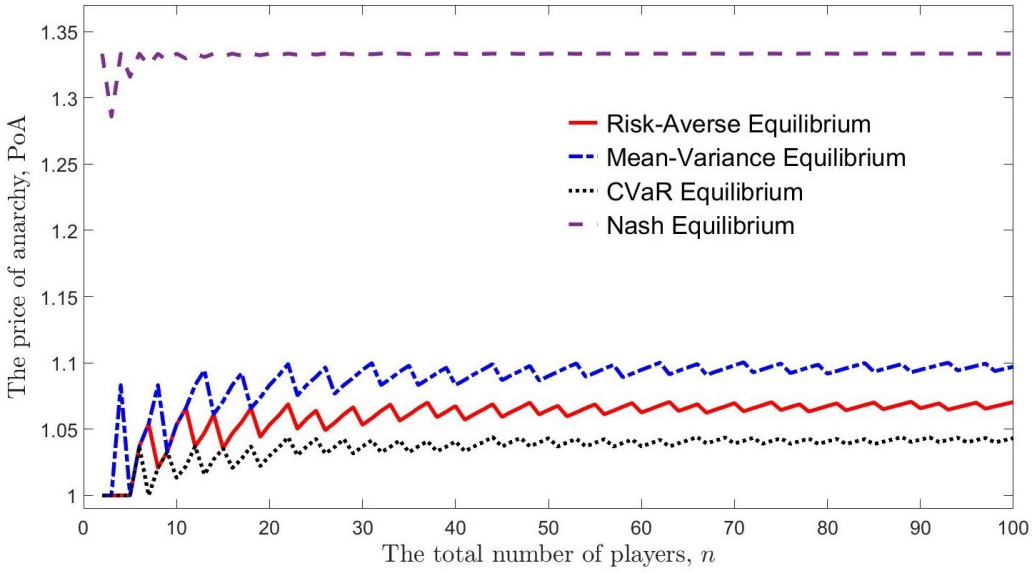

Figure 4.4: The prices of anarchy for the risk-averse, mean-variance ($\rho = 1$), CVaR_α ($\alpha = 0.1$), and Nash equilibria of the Pigou network in Example 6 are plotted for different numbers of players.

close to $\frac{3}{4}$ for $n \gg 1$. If players are risk-neutral and seek to minimize their expected latency given the strategy of the rest of players, which is how the Nash equilibrium models games, the social latency in the mentioned Pigou network equals to one for the Nash equilibrium $(n, 0)$. In contrast, if players are risk-averse in the different senses discussed in this chapter, the social latency decreases compared to when players are risk-neutral; as a result, the price of anarchy decreases as depicted in Figure 4.4. In this example, it is to the benefit of the society if players are risk-averse, which is the case as numerous studies in prospect theory discuss the fact that players in the real world often behave in a risk-averse manner.

Considering the Pigou network in a non-atomic setting, which corresponds to the case with infinite number of players, the socially optimal strategy is $(0.5, 0.5)$ with social latency of $\frac{3}{4}$, where (u_1, u_2) corresponds to u_1 fraction of players traveling along link 1 and $u_2 = 1 - u_1$ fraction of players traveling along link 2. We numerically calculate that the risk-averse equilibrium is $(0.7303, 0.2697)$ with PoA $= 1.0707$, the mean-variance equilibrium with $\rho = 1$ is $(0.7750, 0.2250)$ with PoA $= 1.1008$, the CVaR_α equilibrium with $\alpha = 0.1$ is $(0.6822, 0.3178)$ with PoA $= 1.0442$, and the Nash equilibrium is $(1, 0)$ with PoA $= \frac{4}{3}$.

In the Braess network in Example 7, there are three paths from source to destination, $p_1 = (1, 2)$, $p_2 = (1, 5, 4)$, $p_3 = (3, 4)$, where links SA, AD, SB, BD, and AB are denoted with $1, 2, 3, 4$, and 5, respectively. In order to find the three types of pure equilibria for the Braess network with n players, hypothesize that m^1 players select path p_1, m^2 players select path p_2, and $n - m^1 - m^2$ players select path p_3, then check whether any players has any incentive in the corresponding sense of the equilibrium of the interest to change route, given the pure strategy of the other players. If none of the players has any incentive to change route given the pure strategy of the rest of players, $(m^1, m^2, n - m^1 - m^2)$ is a pure equilibrium. As a result, $(m^1, m^2, n - m^1 - m^2)$ is a pure risk-averse equilibrium if

$$\begin{cases} P\big(L^1 \leq \{L^2, L^3\}\big) \geq \big\{P\big(L^2 \leq \{L^1, L^3\}\big), P\big(L^3 \leq \{L^1, L^2\}\big)\big\}, \text{ where} \\ L^1 = L_1(m^1 + m^2) + L_2(m^1), L^2 = L_1(m^1 + m^2) + L_4(n - m^1 + 1), \text{and} \\ L^3 = L_3(n - m^1 - m^2 + 1) + L_4(n - m^1 + 1), \\ \\ P\big(L^2 \leq \{L^1, L^3\}\big) \geq \big\{P\big(L^1 \leq \{L^2, L^3\}\big), P\big(L^3 \leq \{L^1, L^2\}\big)\big\}, \text{ where} \\ L^1 = L_1(m^1 + m^2) + L_2(m^1 + 1), L^2 = L_1(m^1 + m^2) + L_4(n - m^1), \text{and} \\ L^3 = L_3(n - m^1 - m^2 + 1) + L_4(n - m^1), \\ \\ P\big(L^3 \leq \{L^1, L^2\}\big) \geq \big\{P\big(L^1 \leq \{L^2, L^3\}\big), P\big(L^2 \leq \{L^1, L^3\}\big)\big\}, \text{ where} \\ L^1 = L_1(m^1 + m^2 + 1) + L_2(m^1 + 1), L^2 = L_1(m^1 + m^2 + 1) + L_4(n - m^1), \\ \text{and } L^3 = L_3(n - m^1 - m^2) + L_4(n - m^1). \end{cases} \tag{4.28}$$

By varying m^1 from zero to n and m^2 from 0 to $n - m^1$, if Equation (4.28) holds for $(m^1, m^2, m^3 = n - m^1 - m^2)$, it is a pure risk-averse equilibrium.

Similar to the above approach, $(m^1, m^2, n - m^1 - m^2)$ is a pure mean-

variance equilibrium if

$$\begin{cases} \mathrm{Var}(L^1)+\rho\cdot\mathrm{E}(L^1)\le\big\{\mathrm{Var}(L^2)+\rho\cdot\mathrm{E}(L^2),\mathrm{Var}(L^3)+\rho\cdot\mathrm{E}(L^3)\big\},\ \text{where}\\ L^1=L_1(m^1+m^2)+L_2(m^1),L^2=L_1(m^1+m^2)+L_4(n-m^1+1),\text{and}\\ L^3=L_3(n-m^1-m^2+1)+L_4(n-m^1+1),\\ \\ \mathrm{Var}(L^2)+\rho\cdot\mathrm{E}(L^2)\le\big\{\mathrm{Var}(L^1)+\rho\cdot\mathrm{E}(L^1),\mathrm{Var}(L^3)+\rho\cdot\mathrm{E}(L^3)\big\},\ \text{where}\\ L^1=L_1(m^1+m^2)+L_2(m^1+1),L^2=L_1(m^1+m^2)+L_4(n-m^1),\text{and}\\ L^3=L_3(n-m^1-m^2+1)+L_4(n-m^1),\\ \\ \mathrm{Var}(L^3)+\rho\cdot\mathrm{E}(L^3)\le\big\{\mathrm{Var}(L^1)+\rho\cdot\mathrm{E}(L^1),\mathrm{Var}(L^2)+\rho\cdot\mathrm{E}(L^2)\big\},\ \text{where}\\ L^1=L_1(m^1+m^2+1)+L_2(m^1+1),L^2=L_1(m^1+m^2+1)+L_4(n-m^1),\\ \text{and}\ L^3=L_3(n-m^1-m^2)+L_4(n-m^1). \end{cases} \tag{4.29}$$

By varying m^1 from zero to n and m^2 from 0 to $n-m^1$, if Equation (4.29) holds for $(m^1,m^2,m^3=n-m^1-m^2)$, it is a pure risk-averse equilibrium.

Similar to the above approach, $(m^1,m^2,n-m^1-m^2)$ is a pure CVaR_α equilibrium if

$$\begin{cases} \mathrm{E}\big[L^1\big|L^1\ge v_\alpha^1\big]\le\big\{\mathrm{E}\big[L^2\big|L^2\ge v_\alpha^2\big],\mathrm{E}\big[L^3\big|L^3\ge v_\alpha^3\big]\big\},\ \text{where}\\ L^1=L_1(m^1+m^2)+L_2(m^1),L^2=L_1(m^1+m^2)+L_4(n-m^1+1),\\ L^3=L_3(n-m^1-m^2+1)+L_4(n-m^1+1),\ \text{and}\\ P\big(L^1\ge v_\alpha^1\big)=P\big(L^2\ge v_\alpha^2\big)=P\big(L^3\ge v_\alpha^3\big)=\alpha\\ \\ \mathrm{E}\big[L^2\big|L^2\ge v_\alpha^2\big]\le\big\{\mathrm{E}\big[L^1\big|L^1\ge v_\alpha^1\big],\mathrm{E}\big[L^3\big|L^3\ge v_\alpha^3\big]\big\},\ \text{where}\\ L^1=L_1(m^1+m^2)+L_2(m^1+1),L^2=L_1(m^1+m^2)+L_4(n-m^1),\\ L^3=L_3(n-m^1-m^2+1)+L_4(n-m^1),\ \text{and}\\ P\big(L^1\ge v_\alpha^1\big)=P\big(L^2\ge v_\alpha^2\big)=P\big(L^3\ge v_\alpha^3\big)=\alpha\\ \\ \mathrm{E}\big[L^3\big|L^3\ge v_\alpha^3\big]\le\big\{\mathrm{E}\big[L^1\big|L^1\ge v_\alpha^1\big],\mathrm{E}\big[L^2\big|L^2\ge v_\alpha^2\big]\big\},\ \text{where}\\ L^1=L_1(m^1+m^2+1)+L_2(m^1+1),L^2=L_1(m^1+m^2+1)+L_4(n-m^1),\\ L^3=L_3(n-m^1-m^2)+L_4(n-m^1),\ \text{and}\\ P\big(L^1\ge v_\alpha^1\big)=P\big(L^2\ge v_\alpha^2\big)=P\big(L^3\ge v_\alpha^3\big)=\alpha. \end{cases} \tag{4.30}$$

By varying m^1 from zero to n and m^2 from 0 to $n-m^1$, if Equation (4.30) holds for $(m^1,m^2,m^3=n-m^1-m^2)$, it is a pure CVaR_α equilibrium.

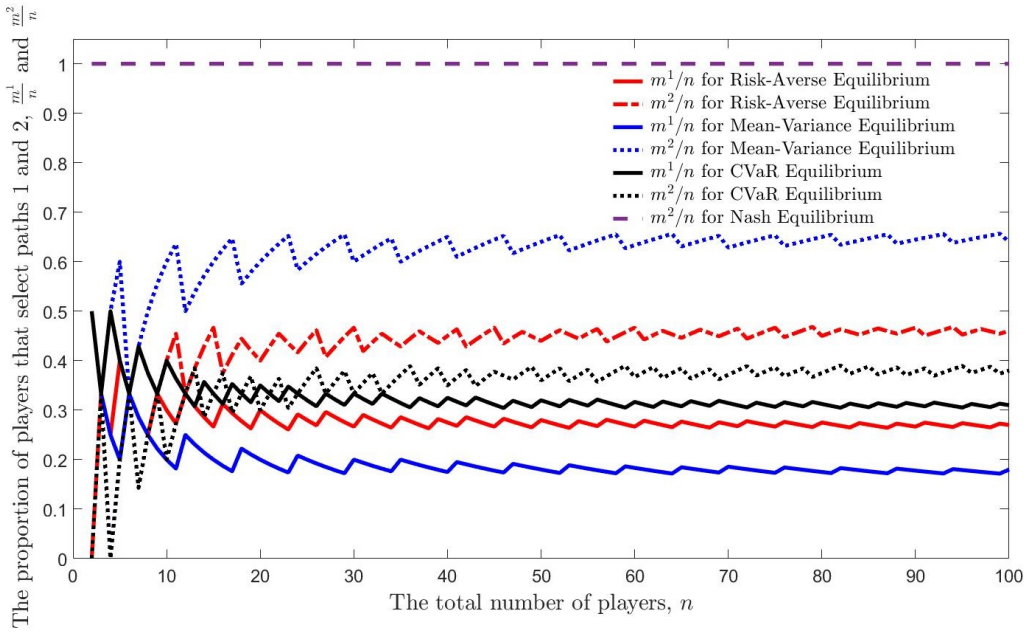

Figure 4.5: The pure risk-averse, mean-variance ($\rho = 1$), CVaR_α ($\alpha = 0.1$), and Nash equilibria of the Braess network in Example 7 are denoted for different numbers of players.

Note that the equilibrium in the Braess network in Example 7 is characterized by m^1 and m^2, since m^3 can be derived given m^1 and m^2. The pure risk-averse, mean-variance ($\rho = 1$), and CVaR_α ($\alpha = 0.1$) equilibria are found for the mentioned Braess network and the proportions of players who select paths 1 and 2, i.e., $\frac{m^1}{n}$ and $\frac{m^2}{n}$, are depicted in Figure 4.5 for different values of n. Under the Nash equilibrium, no matter what the probability distributions of latency over links look like, all players select path 2 as it has less or equal latency in expectation. Hence, $(0, n, 0)$ is the Nash equilibrium for all n, which corresponds to $\frac{m^2}{n} = 1$ and $\frac{m^1}{n} = \frac{m^3}{n} = 0$ as depicted in Figure 4.5.

The social delay/latency defined as the expected average delay/latency incurred by the n players in the Braess network in Example 7 under the pure

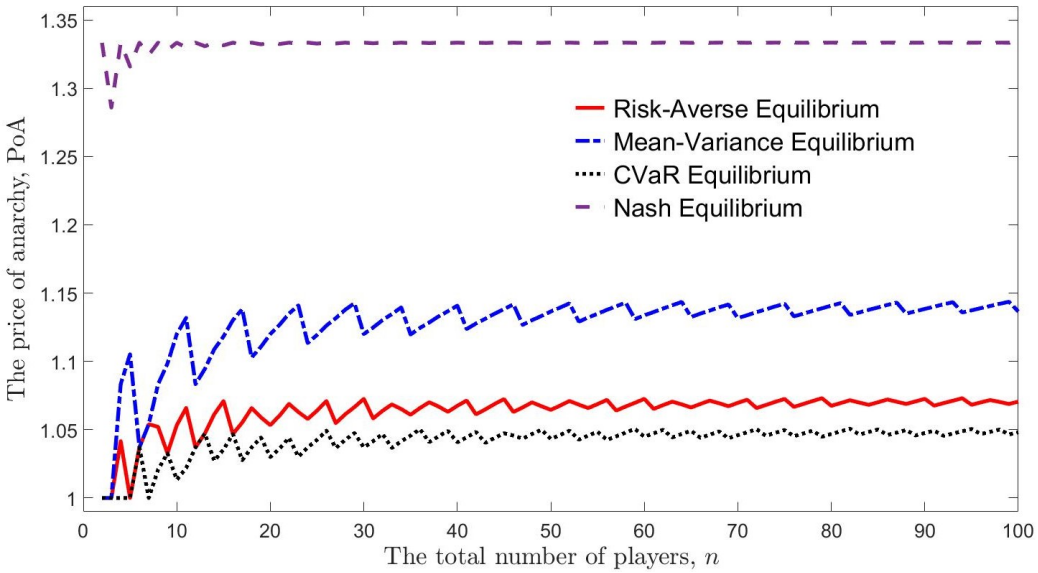

Figure 4.6: The prices of anarchy for the risk-averse, mean-variance ($\rho = 1$), CVaR_α ($\alpha = 0.1$), and Nash equilibria of the Braess network in Example 7 are plotted for different numbers of players.

strategy $(m^1, m^2, m^3 = n - m^1 - m^2)$ is

$$\begin{aligned} D(m^1, m^2) &= \frac{1}{n} \cdot \Bigg((m^1 + m^2) \cdot \frac{(m^1 + m^2)}{n} + m^1 + (n - m^1 - m^2) \\ &\qquad + (n - m^1) \cdot \frac{(n - m^1)}{n} \Bigg) \\ &= \frac{1}{n^2} \cdot \left(2\left(m^1\right)^2 + \left(m^2\right)^2 + 2m^1 m^2 - 2nm^1 - nm^2 + 2n^2 \right), \end{aligned}$$

which is minimized when $\left(m^1 = \lfloor \frac{n}{2} \rfloor, m^2 = 0, m^3 = n - m^1\right)$ or $\left(m^1 = \lceil \frac{n}{2} \rceil, m^2 = 0, m^3 = n - m^1\right)$. As a result, it is socially optimal that about half of players take path p_1 and the rest take path p_3 to travel from source to destination in the Braess network, which results in a social latency close to $\frac{3}{2}$ for $n \gg 1$. If players are risk-neutral and seek to minimize their expected latency given the strategy of the rest of the players, which is how the Nash equilibrium models games, the social latency in the mentioned Braess network equals two for the Nash equilibrium $(0, n, 0)$. In contrast, if players are risk-averse in the different senses discussed in this chapter, the social latency decreases compared to when players are risk-neutral; as a result, the price of anarchy decreases as depicted in Figure 4.6. In this example, it is again to the benefit of the society if players are risk-averse.

Considering the Braess network in a non-atomic setting, which corresponds to the case with infinite number of players, the socially optimal strategy is $(0.5, 0, 0.5)$ with social latency of $\frac{3}{2}$, where (u^1, u^2, u^3) corresponds to u^1 fraction of players travel along path p_1, u^2 fraction of players travel along path p_2, and $u^3 = 1 - u^1 - u^2$ fraction of players travel along path p_3. We numerically calculate that the risk-averse equilibrium is $(0.2655, 0.4690, 0.2655)$ with PoA $= 1.0733$, the mean-variance equilibrium with $\rho = 1$ is $(0.1716, 0.6568, 0.1716)$ with PoA $= 1.1438$, the CVaR_α equilibrium with $\alpha = 0.1$ is $(0.3045, 0.3910, 0.3045)$ with PoA $= 1.0509$, and the Nash equilibrium is $(0, 1, 0)$ with PoA $= \frac{4}{3}$.

Although it is more prevalent to use pure equilibrium for congestion games, we analyze the mixed equilibrium of the Pigou network in Example 6 for two players. The underlying stochastic congestion game with the probability distributions of players' delays, the pure and mixed Nash, risk-averse, mean-variance, and CVaR equilibria are depicted in Figure 4.7. Recall that the (pure) price of anarchy of a congestion game is the maximum ratio $D(\boldsymbol{p})/D(\boldsymbol{o})$ over all equilibria $\boldsymbol{p}$ of the game, where $\boldsymbol{o}$ is the socially optimal strategy. As mentioned earlier, the optimal strategy for the Pigou network with two players is that one of the players travels along the top link and the other player travels along the bottom link which corresponds to the social delay of $\frac{3}{4}$. As a result, the (pure) price of anarchy for the Nash equilibria is $\frac{4}{3}$. On the other hand, the pure price of anarchy for the risk-averse, mean-variance, and CVaR equilibria is equal to one. Furthermore, the price of anarchy among both pure and mixed equilibria for the risk-averse, mean-variance, and CVaR equilibria is 1.2405, 1.1689, and 1.2897, respectively.

In the following, we present extra examples with the purpose of shedding light on drawbacks of the different equilibria in different scenarios and motivating more work to be done on a unified risk-averse framework. Furthermore, the following examples suggest that careful consideration should be given to the choice of the equilibrium that best fits the application of the interest.

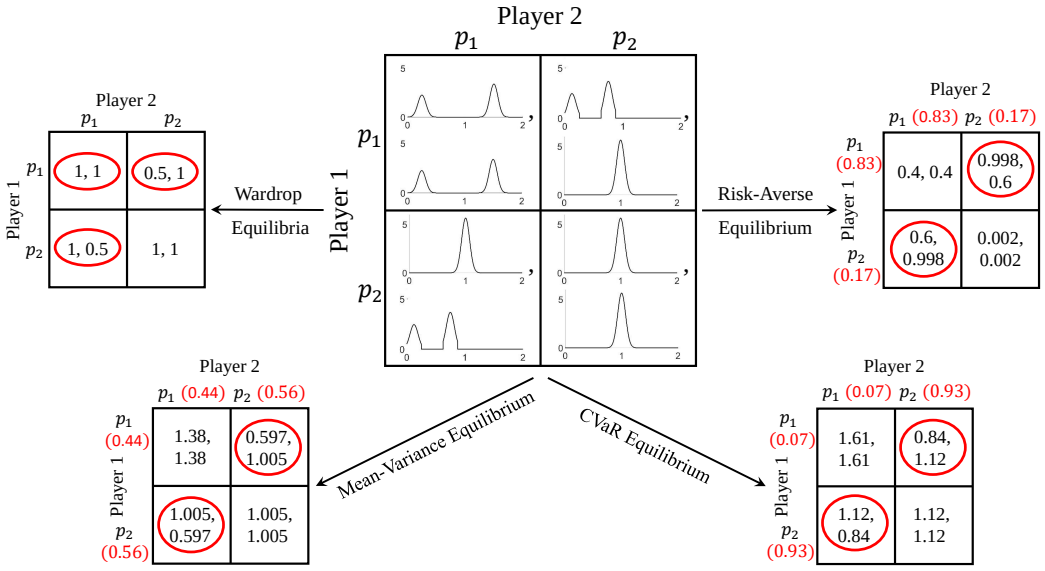

Figure 4.7: The pure and mixed risk-averse, mean-variance ($\rho = 1$), CVaR_α ($\alpha = 0.1$), and Nash equilibria of the Pigou network in Example 6 for two players.

4.3.1 Notes for Practitioners

The intention of this subsection is to direct the attention of practitioners planning to implement risk-averse in-vehicle navigation to cases in which each of the proposed risk-averse equilibria may provide travelers with counterintuitive guidance. To this end, three examples are discussed in the following to shed light on the implications of the three classes of risk-averse equilibria. The examples are meant to be simple to convey the idea in a straightforward manner.

Example 8. *Consider a Pigou network with two parallel links,* 1 *and* 2*, between source and destination. The travel times on links* 1 *and* 2 *are respectively independent random variables* L_1 *and* L_2 *with pdfs*

$$f_1(x) = \alpha\Big(\exp\left(-100\,(x-14)^2\right) \cdot \mathbb{1}\left\{13 \le x \le 15\right\} + \exp\left(-100\,(x-19)^2\right) \cdot \mathbb{1}\left\{18 \le x \le 20\right\} \Big),$$

$$f_2(x) = \beta \exp\left(-100\,(x-20)^2\right) \cdot \mathbb{1}\left\{19 \le x \le 21\right\},$$

where α *and* β *are constants for which each of the two distributions integrate to one.*

In Example 8, the means and variances of travel times along links 1 and 2 are $l_1 = 16.5$, $\mathrm{Var}(L_1) = 6.255$, $l_2 = 20.0$, $\mathrm{Var}(L_2) = 0.005$, respectively, and $P(L_1 \leq L_2) = 1.0$. As a result, although link 1 has a higher variance than link 2, not only is link 1 shorter than link 2 in expectation, but link 1 is shorter than link 2 almost certainly. Hence, a rational traveler intends to take link 1 for commute although its variance is higher than the variance of link 2. However, the mean-variance framework intends to keep a balance between lower expected travel time and lower uncertainty in travel time assuming that higher variance is against the spirit of risk-averse travelers. In Example 8, the mean-variance framework guides travelers to travel along link 2 if $\rho < 1.7857$, which is not optimal from the perspective of a risk-averse traveler. Note that both risk-averse equilibrium and CVaR_α equilibrium for any $\alpha \in [0, 1]$ guide travelers to traverse along link 1 in this example.

Example 9. *Consider a Pigou network with two parallel links,* 1 *and* 2*, between source and destination. The travel times on links* 1 *and* 2 *are respectively independent random variables L_1 and L_2 with pdfs*

$$f_1(x) = \alpha\Big(4\exp\left(-100\left(x-5\right)^2\right)\cdot\mathbb{1}\left\{4\leq x\leq 6\right\} + \exp\left(-100\left(x-10\right)^2\right)\cdot\mathbb{1}\left\{9\leq x\leq 11\right\}\Big),$$

$$f_2(x) = \beta\Big(4\exp\left(-100\left(x-8\right)^2\right)\cdot\mathbb{1}\left\{7\leq x\leq 9\right\} + \exp\left(-100\left(x-10\right)^2\right)\cdot\mathbb{1}\left\{9\leq x\leq 11\right\}\Big),$$

where α and β are constants for which each of the two distributions integrate to one.

In Example 9, the means and variances of travel times along links 1 and 2 are $l_1 = 6.0$, $\mathrm{Var}(L_1) = 4.005$, $l_2 = 8.4$, $\mathrm{Var}(L_2) = 0.645$, respectively, and $P(L_1 \leq L_2) = 0.82$. Note that both distributions are the same over the interval $[9, 11]$; however, the traveler has a better opportunity of experiencing shorter travel time on the lower 0.8 quantile of the distribution of link 1 compared to that of link 2. Hence, a rational traveler intends to take link 1 for commute although its variance is higher than the variance of link 2. Furthermore, $\mathrm{E}\left[L_1 | L_1 \geq \alpha\right] = \mathrm{E}\left[L_2 | L_2 \geq \alpha\right]$ for $\alpha \in [0, 0.2]$; hence, the CVaR_α framework is indifferent between the two links when $\alpha \in [0, 0.2]$, which can

result in a counterintuitive route selection in Example 9. The mean-variance framework also guides travelers to traverse along link 2 if $\rho < 1.4$, which is not optimal from the perspective of a risk-averse traveler. Note that the risk-averse equilibrium guides travelers to traverse along link 1 in this example as $P(L_1 \leq L_2) = 0.82$.

Example 10. *Consider a Pigou network with two parallel links,* 1 *and* 2*, between source and destination. The travel times on links* 1 *and* 2 *are respectively independent random variables L_1 and L_2 with pdfs*

$$f_1(x) = \beta \exp\left(-100\,(x-7)^2\right) \cdot \mathbb{1}\left\{6 \leq x \leq 8\right\},$$
$$f_2(x) = \alpha\Bigg(7\exp\left(-100\,(x-5)^2\right) \cdot \mathbb{1}\left\{4 \leq x \leq 6\right\} + 3\exp\left(-100\,(x-10)^2\right) \cdot \mathbb{1}\left\{9 \leq x \leq 11\right\}\Bigg),$$

where α and β are constants for which each of the two distributions integrate to one.

In Example 10, the means and variances of travel times along links 1 and 2 are $l_1 = 7.0$, $\mathrm{Var}(L_1) = 0.005$, $l_2 = 6.5$, $\mathrm{Var}(L_2) = 5.255$, respectively, and $P(L_2 \leq L_1) = 0.7$. Although the expected travel time along link 2 is less than that along link 1 and it is more likely that the travel time along link 2 is shorter than travel time along link 1, the travel time along link 2 is concentrated around 10 with probability 0.3 which is somewhat larger than the concentration of travel time around 7 when traveling along link 1. Hence, a risk-averse traveler may prefer to take link 1 for commute although its expected travel time is higher than the expected travel time of link 2 to avoid a long travel time. However, the risk-averse equilibrium guides travelers to traverse along link 2, which may not be optimal from the perspective of a risk-averse traveler. Note that the CVaR_α equilibrium for $\alpha < 0.748$ and mean-variance equilibrium for $\rho < 10.5$ guide travelers to traverse along link 1 in this example.

Chapter 5

BLIND GB-PANDAS: A BLIND THROUGHPUT-OPTIMAL LOAD BALANCING ALGORITHM FOR AFFINITY SCHEDULING

Affinity load balancing refers to allocation of computing tasks on computing nodes in an efficient way to minimize a cost function, for example the mean task completion time [199]. Due to the fact that different task types can have different processing (service) rates on different computing nodes (servers), a dilemma between throughput and delay optimality emerges which makes the optimal affinity load balancing an open problem for more than three decades if the task arrival rates are unknown. If the task arrival rates and the service rates of different task types on different servers are known, the fluid model planning algorithm by Harrison and Lopez [120, 121], and Bell and Williams [122, 123], is a delay optimal load balancing algorithm that solves a linear programming optimization problem to determine task assignment on servers. The same number of queues as the number of task types is needed for the fluid model planning algorithm, so the queueing structure is fixed to the number of task types and does not capture the complexity of the system model, which is how heterogeneous the service rates of task types on different servers are. As an example given in [134] and [201], for data centers with a rack structure that use Hadoop for MapReduce data placement with three replicas of data chunks on the M severs, the fluid model planning algorithm requires $\binom{M}{3}$ queues, while Xie et al. [134] propose a delay optimal algorithm that uses $3M$ queues. As another extreme example, if the service rates of N_T number of task types on all servers are the same, the fluid model planning algorithm still considers N_T number of queues, while the First-Come-First-Served (FCFS) algorithm uses a single queue and is both throughput and delay optimal. It is true that in the last example all task types can be considered the same type, but this is just an example to illuminate the reasoning behind the queueing structure

Portions of this chapter were previously published in Yekkehkhany and Nagi [131] and are used here with permission. Furthermore, portions of this chapter were previously published in Yekkehkhany et al. [200] and are used here with permission.

for GB-PANDAS (Generalized Balanced Priority Algorithm for Near Data Scheduling) presented in Subsection 5.4.1.

In the absence of knowledge on task arrival rates, Max-Weight [127] and c-μ-rule [124] algorithms can stabilize the system by just knowing the service rates of task types on different servers. None of these two algorithms are delay optimal though. The c-μ-rule is actually cost optimal, where it assumes convex delay costs associated with each task type, and minimizes the total cost incurred by the system. Since the cost functions have to be strictly convex, and so cannot be linear, c-μ-rule does not minimize the mean task completion time. Since these two algorithms do not use the task arrival rates and still stabilize the system, they are robust to any changes in task arrival rate as long as it is in the capacity region of the system. Both Max-Weight and c-μ-rule algorithms have the same issue as the fluid model planning algorithm on considering one queue per task type which can make the system model complicated as discussed in [134]. Note that Wang et al. [132] and Xie et al. [134] study the load balancing problem for special cases of two and three levels of data locality, respectively. In the former, delay optimality is analyzed for a special traffic scenario and in the latter delay optimality is analyzed for a general traffic scenario. In both cases there is no issue with the number of queues, but as mentioned, these two algorithms are for special cases of two and three levels of data locality. Hence, a unified algorithm that captures the trade-off between the complexity of the queueing structure and the complexity of the system model is lacking in the literature. Yekkehkhany et al. [200] implicitly mention this trade-off in data center applications, but the generalization is not crystal clear and needs more consideration of the affinity setup, which is summarized in this work as a complementary note on the Balanced-PANDAS algorithm.

The affinity scheduling problem appears in different applications from data centers and modern processing networks that consist of heterogeneous servers, where data-intensive analytics like MapReduce, Hadoop, and Dryad are performed, to supermarket models, or even patient assignment to surgeons in big and busy hospitals and many more. Lack of dependable estimates of system parameters, including task arrival rates and specially service rates of task types on different servers, is a major challenge in constructing an optimal load balancing algorithm for such networks [202]. All the algorithms mentioned above at least require the knowledge of service rates of task

types on different servers. In the absence of prior knowledge on service rates, such algorithms can be fragile and perform poorly, resulting in huge waste of resources. To address this issue, we propose a robust policy called Blind GB-PANDAS that is totally blind to all system parameters, but is robust to task arrival rate changes and learns the service rates of task types on different servers, so it is robust to any service rate parameter changes as well. It is natural that due to traffic load changes in data centers, the service rates of tasks on remote servers change over time. In such cases, Blind GB-PANDAS is capable of updating system parameters and taking action correspondingly. Blind GB-PANDAS uses an exploration-exploitation approach to make the system stable without any knowledge about the task arrival rates and the processing rates. More specifically, it uses an exploration-exploitation method, where in the exploration phase it takes action in a way to make the system parameter estimations more accurate, and in the exploitation phase it uses the estimated parameters to do an optimal load balancing based on the estimates. Note that only the processing rates of task types on different servers are the parameters that are estimated, and the task arrival rates are not estimated. The reason is that task arrival rates change frequently, so there is no point estimating them, whereas the service rates do not change rapidly. Since Blind GB-PANDAS uses an estimate of the processing rates, an incoming task is not necessarily routed to the server with the minimum weighted-workload in the exploitation phase, which raises the complexity in the throughput optimality proof of Blind GB-PANDAS using the Lyapunov-based method. The throughput optimality result is proved under arbitrary and unknown service time distributions with bounded means and bounded supports that do not necessarily require the memoryless property.

As discussed in Subsection 5.4.1, the queueing structure used for Blind GB-PANDAS shows the trade-off between the heterogeneity of the underlying system model for processing rates and the complexity of the Blind GB-PANDAS queueing structure. Blind GB-PANDAS can also use a one-queue-per-server queueing structure, where the workload on servers is of interest instead of the queue lengths, but for an easier explanation of the Blind GB-PANDAS algorithm we use multiple symbolic sub-queues for each server. The Blind GB-PANDAS algorithm is compared to FCFS, Max-Weight, and c-μ-rule algorithms in terms of average task completion time through simulations, where the same exploration-exploitation approach as Blind GB-PANDAS is

used for Max-Weight and c-μ-rule. Our extensive simulations show that the Blind GB-PANDAS algorithm outperforms the two other algorithms at high loads by an obviously large difference.

The rest of the chapter is structured as follows. Section 5.1 describes the system model for data centers with a nested rack structure, Section 5.2 presents the GB-PANDAS algorithm for such a system, and Section 5.3 provides the throughput optimality proof for the GB-PANDAS algorithm. Section 5.4 describes the system model, GB-PANDAS, and the queueing structure of GB-PANDAS, in addition to deriving the capacity region of the system. Section 5.5 presents the Blind GB-PANDAS algorithm and queueing dynamics for this algorithm. Section 5.6 starts with some preliminary results and lemmas and ends up with the throughput optimality proof for Blind GB-PANDAS. Section 5.7 evaluates the performance of Blind GB-PANDAS versus Max-Weight, c-μ-rule, and FCFS algorithms in terms of mean task completion time. For the conclusion of this chapter and a discussion of opportunities for future work, refer to Chapter 6.

5.1 Data Centers with a Nested Rack Structure

In this section, we propose the Generalized-Balanced-Priority-Algorithm-for-Near-Data-Scheduling (Generalized-Balanced-Pandas or GB-PANDAS) with a new queueing structure for a data center with a nested rack structure as described later. The GB-PANDAS algorithm does not require the arrival rates of task types and is for a case with multiple levels of data localities. We establish the capacity region of the system with a nested rack structure and prove the throughput optimality of the proposed algorithm. The service times are assumed to be non-preemptive and they can have an arbitrary distribution, not necessarily geometric distribution which is the main assumption in [133, 134], so we have to use a different Lyapunov function than the ordinary sum of cubic of the queue lengths to prove the throughput optimality of the GB-PANDAS algorithm. We take the map task scheduling problem, which is described in Subsection 1.2.5, as a platform to test the performance of the proposed algorithm versus the state-of-the-art algorithms that are either widely used in the industry or have theoretical guarantees for optimality in some senses. The extensive simulation results show that the GB-PANDAS

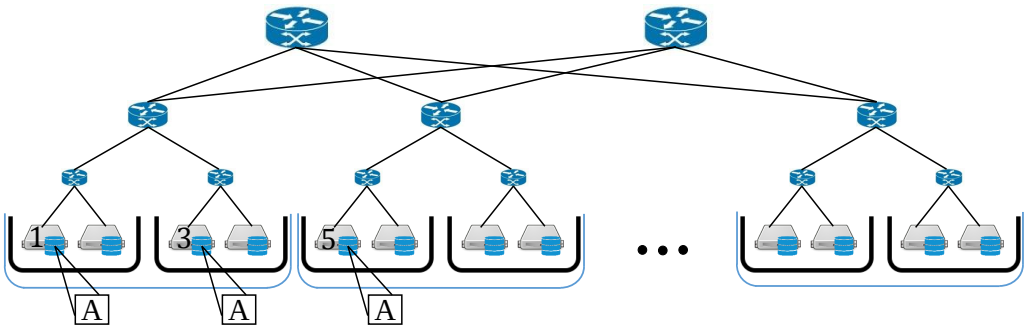

Figure 5.1: A typical data center architecture with four levels of data locality.

algorithm performs better than other algorithms at heavy-traffic loads.

5.1.1 The System Model for a Data Center with a Nested Rack Structure

A discrete time model for the system is studied, where time is indexed by $t \in \mathbb{N}$. The system consists of M servers indexed by $1, 2, \cdots, M$. Let $\mathcal{M} = \{1, 2, \cdots, M\}$ be the set of servers. In today's typical data center architecture, these servers are connected to each other through different levels of switches or routers. A typical data center architecture is shown in Figure 5.1, which consists of servers, racks, super racks, top of the rack switches, top of the super rack switches, and core switches.

Remark 8. *Note that our theoretical analysis does not care about the rack structure in data centers, so the result of throughput optimality of the GB-PANDAS algorithm is proved for an arbitrary system with N levels of data locality (as an example, recall the affinity scheduling problem). The rack structure is only proposed as an incentive for this theoretical work, but the result is more general.*

Considering the MapReduce framework for processing large data-sets, the data-set is split into small data chunks (typically of size 128 MB), and the data chunks are replicated on d servers where the default for Hadoop is $d = 3$ servers. The bottleneck in MapReduce is the Map tasks, not the Reduce task, so we only consider Map tasks in this chapter.

Task Type: In the Map stage, each task is associated with the processing of a data chunk, and by convention we denote the type of the task by the

label of the three servers where the data chunk is stored [106, 134]. As an example, the task associated with processing data chunk A shown in Figure 5.1 has type $\bar{L} = (1, 3, 5)$ since data chunk A is stored in these three servers. The set of all task types $\bar{L}$ is denoted by $\mathcal{L}$ defined as follows:

$$\bar{L} \in \mathcal{L} = \{(m_1, m_2, m_3) \in \mathcal{M}^3 : m_1 < m_2 < m_3\},$$

where m_1, m_2, and m_3 are the three local servers.[1] A task of type $\bar{L} = (m_1, m_2, m_3)$ receives faster average service from its local servers than from servers that do not have the data chunk. The reason is that the server without the data chunk has to fetch data associated to a task of type $\bar{L}$ from any of its local servers. According to the distance between the two servers, this fetching of the data can cause different amounts of delay. This fact brings the different levels of data locality into account. Obviously, the closer the two servers, the shorter the delay. Hence, the communication cost through the network and switches between two servers in the same rack is less than that between two servers in the same super rack (but different racks), and the cost for both is on average less than that between two servers in different super racks. Generally speaking, we propose the N levels of data locality as follows:

Service Process: The non-preemptive service (processing) time of a task of type $\bar{L} = (m_1, m_2, m_3) \in \mathcal{L}$ is a random variable with cumulative distribution function (CDF)

- F_1 with mean $\frac{1}{\alpha_1}$ if the task receives service from any server in the set $\bar{L} = \{m_1, m_2, m_3\}$, and we say that the task is 1-local to these servers.
- F_n with mean $\frac{1}{\alpha_n}$ if the task receives service from any server in the set $\bar{L}_n$, defined in the following, and we say that the task is n-local to these servers, for $n \in \{2, 3, \cdots, N\}$,

where $\alpha_1 > \alpha_2 > \cdots > \alpha_N$.

In the data center structure example in Figure 5.1, the set $\bar{L}_2$ is the set of all servers that do not have the data saved on their own disk, but data is stored in another server in the same rack; and the set $\bar{L}_3$ is the set of all servers that

[1] The analysis is not sensitive to the number of local servers. The default number of local servers in Hadoop is three, so we choose three local servers, but this assumption can be ignored without any change in the analysis.

do not have the data saved on their own disk, but data is stored in another server in another rack, but in the same super rack, and so on.

Remark 9. *Note that the service time is not necessarily assumed to be geometrically distributed and can be arbitrary as long as it satisfies the decreasing property of the means mentioned above.*

Arrival Process: The number of arriving tasks of type $\bar{L}$ at the beginning of time slot t is denoted by $A_{\bar{L}}(t)$, which are assumed to be temporarily i.i.d. with mean $\lambda_{\bar{L}}$. The total number of arriving tasks at each time slot is assumed to be bounded by a constant C_A and is assumed to be zero with a positive probability. The set of all arrival rates for different types of tasks is denoted by the vector $\boldsymbol{\lambda} = (\lambda_{\bar{L}} : \bar{L} \in \mathcal{L})$.

5.1.2 An Outer Bound of the Capacity Region for a Data Center with a Nested Rack Structure

The arrival rate of type $\bar{L}$ tasks can be decomposed to $(\lambda_{\bar{L},m}, m \in \mathcal{M})$, where $\lambda_{\bar{L},m}$ denotes the arrival rate of type $\bar{L}$ tasks that are processed by server m. Obviously, $\sum_{m\in\mathcal{M}} \lambda_{\bar{L},m} = \lambda_{\bar{L}}$. A necessary condition for an arrival rate vector $\boldsymbol{\lambda}$ to be supportable is that the total 1-local, 2-local, $\cdots$, N-local load on each server be strictly less than one for all servers as the following inequality suggests:

$$\sum_{\bar{L}:m\in\bar{L}} \frac{\lambda_{\bar{L},m}}{\alpha_1} + \sum_{\bar{L}:m\in\bar{L}_2} \frac{\lambda_{\bar{L},m}}{\alpha_2} + \cdots + \sum_{\bar{L}:m\in\bar{L}_N} \frac{\lambda_{\bar{L},m}}{\alpha_N} < 1,\ \forall m \in \mathcal{M}. \tag{5.1}$$

Given this necessary condition, an outer bound of the capacity region is given by the set of all arrival rate vectors $\boldsymbol{\lambda}$ with a decomposition satisfying (5.1) as follows.

$$\begin{aligned} \Lambda = \Big\{ &\boldsymbol{\lambda} = (\lambda_{\bar{L}} : \bar{L} \in \mathcal{L}) \ \Big|\ \exists \lambda_{\bar{L},m} \geq 0, \forall \bar{L} \in \mathcal{L}, \forall m \in \mathcal{M}, s.t. \\ &\lambda_{\bar{L}} = \sum_{m=1}^{M} \lambda_{\bar{L},m},\ \forall \bar{L} \in \mathcal{L}, \\ &\sum_{\bar{L}:m\in\bar{L}} \frac{\lambda_{\bar{L},m}}{\alpha_1} + \sum_{\bar{L}:m\in\bar{L}_2} \frac{\lambda_{\bar{L},m}}{\alpha_2} + \cdots + \sum_{\bar{L}:m\in\bar{L}_N} \frac{\lambda_{\bar{L},m}}{\alpha_N} < 1, \forall m \Big\}. \end{aligned} \tag{5.2}$$

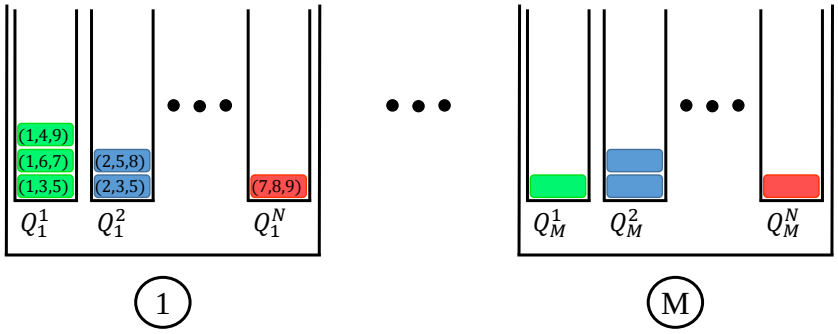

Figure 5.2: The queueing structure when the GB-PANDAS algorithm is used.

It is clear that to find Λ, we should solve a linear programming optimization problem. We will show in Section 5.3 that GB-PANDAS stabilizes the system as long as the arrival rate vector $\boldsymbol{\lambda}$ is inside Λ, which means that this outer bound of the capacity region is the capacity region itself. In the following, Lemma 1 proposes a set which is equivalent to that in (5.2) which will be used in the throughput optimality proof of GB-PANDAS.

Lemma 1. *The following set $\bar{\Lambda}$ is equivalent to Λ defined in equation* (5.2)*:*

$$\bar{\Lambda} = \Bigg\{ \boldsymbol{\lambda} = (\lambda_{\bar{L}} : \bar{L} \in \mathcal{L}) \Big| \exists \lambda_{\bar{L},n,m} \geq 0, \forall \bar{L} \in \mathcal{L}, \forall n \in \bar{L}, \forall m \in \mathcal{M}, s.t.$$

$$\lambda_{\bar{L}} = \sum_{n:n\in\bar{L}} \sum_{m=1}^{M} \lambda_{\bar{L},n,m}, \ \forall \bar{L} \in \mathcal{L},$$

$$\sum_{\bar{L}:m\in\bar{L}} \sum_{n:n\in\bar{L}} \frac{\lambda_{\bar{L},n,m}}{\alpha_1} + \sum_{\bar{L}:m\in\bar{L}_2} \sum_{n:n\in\bar{L}} \frac{\lambda_{\bar{L},n,m}}{\alpha_2} +$$

$$\cdots + \sum_{\bar{L}:m\in\bar{L}_N} \sum_{n:n\in\bar{L}} \frac{\lambda_{\bar{L},n,m}}{\alpha_N} < 1, \forall m \Bigg\}, \tag{5.3}$$

where $\lambda_{\bar{L},n,m}$ denotes the arrival rate of type $\bar{L}$ tasks that are 1-local to server n and is processed by server m. $\{\lambda_{\bar{L},n,m} : \bar{L} \in \mathcal{L}, n \in \bar{L}, \text{and } m \in \mathcal{M}\}$ is a decomposition of the set of arrival rates $\{\lambda_{\bar{L},m} : \bar{L} \in \mathcal{L} \text{ and } m \in \mathcal{M}\}$, where $\lambda_{\bar{L},m} = \sum_{n\in\mathcal{M}} \lambda_{\bar{L},n,m}$.

The proof of Lemma 1 is provided in Appendix D.1.

5.2 The GB-PANDAS Algorithm for a Data Center with a Nested Rack Structure

The central scheduler keeps N queues per server as shown in Figure 5.2. The N queues of the m-th server are denoted by $Q_m^1, Q_m^2, \cdots, Q_m^N$. Tasks that are routed to server m and are n-local to this server are queued at queue Q_m^n. The length of this queue, defined as the number of tasks queued in this queue, at time slot t, is shown by $Q_m^n(t)$. The central scheduler maintains the length of all queues at all time slots, which is denoted by vector $\boldsymbol{Q}(t) = \big(Q_1^1(t), Q_1^2(t), \cdots, Q_1^N(t), \cdots, Q_M^1(t), Q_M^2(t), \cdots, Q_M^N(t)\big)$. In the following, the workload on a server is defined which will be used in the statement of the GB-PANDAS algorithm.

Workload of Server m: Under the GB-PANDAS algorithm, server m only processes tasks that are queued in its N queues, that is $Q_m^1, Q_m^2, \cdots, Q_m^N$. As the processing time of an n-local task follows a distribution with CDF F_n and mean $\frac{1}{\alpha_n}$, the expected time needed for server m to process all tasks queued in its queues at time slot t is given as follows:

$$W_m(t) = \frac{Q_m^1(t)}{\alpha_1} + \frac{Q_m^2(t)}{\alpha_2} + \cdots + \frac{Q_m^N(t)}{\alpha_N}.$$

We name $W_m(t)$ the *workload* on the m-th server.

A load balancing algorithm consists of two parts, routing and scheduling. The routing policy determines the queue at which a new incoming task is queued until it receives service from a server. When a server becomes idle and so is ready to process another task, the scheduling policy determines the task receiving service from the idle server. The routing and scheduling policies of the GB-PANDAS algorithm are as follows:

- **GB-PANDAS Routing (Weighted-Workload Routing):** The incoming task of type $\bar{L}$ is routed to the corresponding sub-queue of server m^* with the minimum weighted workload as defined in the following (ties are broken randomly):

  $$m^* = \arg\min_{m \in \mathcal{M}} \left\{ \frac{W_m(t)}{\alpha_1} I_{\{m \in \bar{L}\}} + \frac{W_m(t)}{\alpha_2} I_{\{m \in \bar{L}_2\}} + \cdots + \frac{W_m(t)}{\alpha_N} I_{\{m \in \bar{L}_N\}} \right\}.$$

  If this task of type $\bar{L}$ is 1-local, 2-local, $\cdots$, N-local to server m^*, it is queued at $Q_{m^*}^1$, $Q_{m^*}^2$, $\cdots$, $Q_{m^*}^N$, respectively.

- **GB-PANDAS Scheduling (Prioritized Scheduling):** The idle server m is only scheduled to process a task from its own queues, Q_m^1, Q_m^2, $\cdots$, Q_m^N. A task that is n-local to server m has a higher priority than a task that is $(n+1)$-local to server m (for $1 \leq n \leq N-1$). Hence, the idle server m keeps processing a task from Q_m^1 until there are no more tasks available at this queue, then continues processing tasks queued at Q_m^2, and so on.

5.2.1 The Queueing Dynamics for a Data Center with a Nested Rack Structure

Denote the number of arriving tasks at Q_m^n at time slot t by $A_m^n(t)$, where these tasks are n-local to server m. Recall the notation $A_{\bar{L},m}(t)$ for the number of tasks of type $\bar{L}$ that are scheduled to server m. Then, we have the following relation between $A_m^n(t)$ and $A_{\bar{L},m}(t)$:

$$\begin{aligned} A_m^1(t) &= \sum_{\bar{L}:m\in\bar{L}} A_{\bar{L},m}(t), \\ A_m^n(t) &= \sum_{\bar{L}:m\in\bar{L}_n} A_{\bar{L},m}(t), \quad \text{for } \ 2 \leq n \leq N, \end{aligned} \tag{5.4}$$

where $\bar{L}$ is the set of 1-local servers and $\bar{L}_n$ for $2 \leq n \leq N$ is the set of n-local servers to a task of type $\bar{L}$. The number of tasks that receive service from server m at time slot t and are n-local to the server is denoted by $S_m^n(t)$ which is the number of departures from Q_m^n (as a reminder, the service time of a task that is n-local to a server has CDF F_n). Then, the queue dynamics for any $m \in \mathcal{M}$ are as follows:

$$\begin{aligned} Q_m^n(t+1) &= Q_m^n(t) + A_m^n(t) - S_m^n(t), \quad \text{for } \ 1 \leq n \leq N-1, \\ Q_m^N(t+1) &= Q_m^N(t) + A_m^N(t) - S_m^N(t) + U_m(t), \end{aligned} \tag{5.5}$$

where $U_m(t) = \max\left\{0, S_m^N(t) - A_m^N(t) - Q_m^N(t)\right\}$ is the unused service of server m.

Note that the set of queue lengths $\{\boldsymbol{Q}(t), t \geq 0\}$ do not form a Markov chain since not having the information about how long a server has been processing a task and what type that task is, leads to $\boldsymbol{Q}(t+1)|\boldsymbol{Q}(t) \not\perp \boldsymbol{Q}(t-1)|\boldsymbol{Q}(t)$. Note that the processing time of a task has a general CDF, not necessarily geometric distribution with memoryless property, so we do

need to consider two parameters about the status of servers in the system as follows to be able to define a Markov chain.

- Let $\Psi_m(t)$ be the number of time slots at the beginning of time slot t that server m has spent on the currently in-service task. Note that $\Psi_m(t)$ is set to zero when server m is done processing a task. Then the first working status vector, $\boldsymbol{\Psi}(t)$, is defined as follows:

$$\boldsymbol{\Psi}(t) = \big(\Psi_1(t), \Psi_2(t), \cdots, \Psi_M(t)\big).$$

- The second working status vector is $\boldsymbol{f}(t) = \big(f_1(t),\, f_2(t), \cdots,\, f_M(t)\big)$, where

$$f_m(t) = \begin{cases} -1, \text{ if server } m \text{ is idle,} \\ 1, \text{ if server } m \text{ processes a 1-local task from } Q_m^1, \\ 2, \text{ if server } m \text{ processes a 2-local task from } Q_m^2, \\ \vdots \\ N, \text{ if server } m \text{ processes an N-local task from } Q_m^N. \end{cases}$$

Define $\eta_m(t)$ as the scheduling decision for server m at time slot t. If server m finishes the processing of an in-service task at time slot t, we have $f_m(t^-) = -1$ and the central scheduler makes the scheduling decision $\eta_m(t)$ for the idle server m. Note that $\eta_m(t) = f_m(t)$ as long as server m is processing a task. Then, we define the following vector:

$$\boldsymbol{\eta}(t) = \big(\eta_1(t), \eta_2(t), \cdots, \eta_M(t)\big).$$

As mentioned, since the service times have a general distribution with arbitrary CDF but not necessarily geometrically distributed, the queueing process — or even both the queueing and $\boldsymbol{\eta}(t)$ processes — do not form a Markov chain (one reason is that the service time does not have the memoryless property). Therefore, we consider the Markov chain $\big\{\boldsymbol{Z}(t) = \big(\boldsymbol{Q}(t), \boldsymbol{\eta}(t), \boldsymbol{\Psi}(t)\big), t \geq 0\big\}$ and show that it is irreducible and aperiodic. The state space of this Markov chain is $\mathcal{S} = \mathbb{N}^{NM} \times \{1, 2, \cdots, N\}^M \times \mathbb{N}^M$. Assume the initial state of the Markov chain to be $\boldsymbol{Z}(0) = \big\{0_{NM\times 1}, N_{M\times 1}, 0_{M\times 1}\big\}$.
Irreducible: Since the CDF of the service times, F_n for $1 \leq n \leq N$, are increasing, there exists a positive integer τ such that $F_n(\tau) > 0$ for $1 \leq n \leq N$.

Moreover, the probability of zero arrival tasks is positive. Hence, for any state of the system, $\boldsymbol{Z} = \big(\boldsymbol{Q}, \boldsymbol{\eta}, \boldsymbol{\Psi}\big)$, the probability of the event that each job gets processed in τ time slots and no tasks arrive at the system in $\tau \sum_{m=1}^{M} \sum_{n=1}^{N} Q_m^n$ time slots is positive. As a result, the initial state is reachable from any state in the state space and $\big\{\boldsymbol{Z}(t)\big\}$ is irreducible.
Aperiodic: Since the probability of zero arriving tasks is positive, there is a positive probability of transition from the initial state to itself. Then, given that $\big\{\boldsymbol{Z}(t)\big\}$ is irreducible, it is also aperiodic.

5.3 Throughput Optimality of the GB-PANDAS Algorithm for a Data Center with a Nested Rack Structure

Theorem 8. *The GB-PANDAS algorithm stabilizes a system with N levels of data locality as long as the arrival rate is strictly inside the capacity region, which means that the Generalized Balanced-Pandas algorithm is throughput optimal.*

Proof. The throughput optimality proof of the GB-PANDAS algorithm for a system with N levels of data locality and a general service time distribution follows an extension of the Foster-Lyapunov theorem as stated below.
Extended Version of the Foster-Lyapunov Theorem (Theorem 3.3.8 in [203]): Consider an irreducible Markov chain $\{Z(t)\}$, where $t \in \mathbb{N}$, with a state space $\mathcal{S}$. If there exists a function $V : \mathcal{S} \to \mathcal{R}^+$, a positive integer $T \geq 1$, and a finite set $\mathcal{P} \subseteq \mathcal{S}$ satisfying the following condition:

$$\begin{aligned} &\mathbb{E}\left[V(Z(t_0+T)) - V(Z(t_0))|Z(t_0) = z\right] \\ \leq &-\theta \mathbb{I}_{\{z \in \mathcal{P}^c\}} + C\mathbb{I}_{\{z \in \mathcal{P}\}}, \end{aligned} \tag{5.6}$$

for some $\theta > 0$ and $C < \infty$, then the irreducible Markov chain $\{Z(t)\}$ is positive recurrent.

Consider the Markov chain $\big\{\boldsymbol{Z}(t) = \big(\boldsymbol{Q}(t), \boldsymbol{\eta}(t), \boldsymbol{\Psi}(t)\big), t \geq 0\big\}$. As long as the arrival rate vector is strictly inside the outer bound of the capacity region, $\boldsymbol{\lambda} \in \Lambda$, and using the GB-PANDAS algorithm, if we can prove that this Markov chain is positive recurrent, the distribution of $Z(t)$ converges to its stationary distribution when $t \to \infty$, which results in the stability of the

system, so the throughput optimality of the GB-PANDAS algorithm will be proved.

As shown before, the Markov chain $Z(t)$ is irreducible and aperiodic for any arrival rate vector strictly inside the outer bound of the capacity region, $\boldsymbol{\lambda} \in \Lambda$. Hence, if we can find a Lyapunov function $V(.)$ satisfying the drift condition in the extended version of the Foster-Lyapunov theorem when using the GB-PANDAS algorithm, the stability of the system under this algorithm is proved. Lemmas 2, 3, and 4 followed by our choice of the Lyapunov function presented afterwards complete the proof.

Since Λ is an open set, for any $\boldsymbol{\lambda} \in \Lambda$ there exists $\delta > 0$ such that $\boldsymbol{\lambda}' = (1+\delta)\boldsymbol{\lambda} \in \Lambda$ which means that $\boldsymbol{\lambda}'$ satisfies the conditions in (5.2) and specifically the inequality (5.1). Then we have the following for any $m \in \mathcal{M}$:

$$\sum_{\bar{L}:m\in\bar{L}} \frac{\lambda_{\bar{L},m}}{\alpha_1} + \sum_{\bar{L}:m\in\bar{L}_2} \frac{\lambda_{\bar{L},m}}{\alpha_2} + \cdots + \sum_{\bar{L}:m\in\bar{L}_N} \frac{\lambda_{\bar{L},m}}{\alpha_N} < \frac{1}{1+\delta}. \tag{5.7}$$

The load decomposition $\{\lambda_{\bar{L},m}\}$ can be interpreted as one possibility of assigning the arrival rates to the M servers so that the system becomes stable. We then define the ideal workload on each server m under the load decomposition $\{\lambda_{\bar{L},m}\}$ as

$$w_m = \sum_{\bar{L}:m\in\bar{L}} \frac{\lambda_{\bar{L},m}}{\alpha_1} + \sum_{\bar{L}:m\in\bar{L}_2} \frac{\lambda_{\bar{L},m}}{\alpha_2} + \cdots + \sum_{\bar{L}:m\in\bar{L}_N} \frac{\lambda_{\bar{L},m}}{\alpha_N},\ \forall m \in \mathcal{M}. \tag{5.8}$$

Let $\boldsymbol{w} = (w_1, w_2, \cdots, w_M)$, where Lemmas 3 and 4 use this ideal workload on servers as an intermediary to later prove the throughput optimality of the GB-PANDAS algorithm.

The dynamic of the workload on server m, $W_m(.)$, is as follows:

$$
\begin{aligned}
W_m(t+1) &= \frac{Q_m^1(t+1)}{\alpha_1} + \frac{Q_m^2(t+1)}{\alpha_2} + \cdots + \frac{Q_m^N(t+1)}{\alpha_N} \\
&\overset{(a)}{=} \frac{Q_m^1(t) + A_m^1(t) - S_m^1(t)}{\alpha_1} + \frac{Q_m^2(t) + A_m^2(t) - S_m^2(t)}{\alpha_2} + \\
&\qquad \cdots + \frac{Q_m^N(t) + A_m^N(t) - S_m^N(t) + U_m(t)}{\alpha_N} \\
&= W_m(t) + \left(\frac{A_m^1(t)}{\alpha_1} + \frac{A_m^2(t)}{\alpha_2} + \cdots + \frac{A_m^N(t)}{\alpha_N} \right) \\
&\qquad - \left(\frac{S_m^1(t)}{\alpha_1} + \frac{S_m^2(t)}{\alpha_2} + \cdots + \frac{S_m^N(t)}{\alpha_N} \right) + \frac{U_m(t)}{\alpha_N} \\
&\overset{(b)}{=} W_m(t) + A_m(t) - S_m(t) + \widetilde{U}_m(t),
\end{aligned}
$$

where (a) follows from the queue dynamic in (5.5) and (b) is true by the following definitions:

$$
\begin{aligned}
A_m(t) &= \frac{A_m^1(t)}{\alpha_1} + \frac{A_m^2(t)}{\alpha_2} + \cdots + \frac{A_m^N(t)}{\alpha_N}, \ \forall m \in \mathcal{M}, \\
S_m(t) &= \frac{S_m^1(t)}{\alpha_1} + \frac{S_m^2(t)}{\alpha_2} + \cdots + \frac{S_m^N(t)}{\alpha_N}, \ \forall m \in \mathcal{M}, \\
\widetilde{U}_m(t) &= \frac{U_m(t)}{\alpha_N}, \ \forall m \in \mathcal{M}.
\end{aligned} \tag{5.9}
$$

$\boldsymbol{A} = (A_1, A_2, \cdots, A_M)$, $\boldsymbol{S} = (S_1, S_2, \cdots, S_M)$, and $\widetilde{\boldsymbol{U}} = (\widetilde{U}_1, \widetilde{U}_2, \cdots, \widetilde{U}_M)$ are the pseudo task arrival, service and unused service processes, respectively.

The workload on servers, which is denoted by $\boldsymbol{W} = (W_1, W_2, \cdots, W_M)$, has the following dynamic

$$
\boldsymbol{W}(t+1) = \boldsymbol{W}(t) + \boldsymbol{A}(t) - \boldsymbol{S}(t) + \widetilde{\boldsymbol{U}}(t). \tag{5.10}
$$

Lemmas 2, 3, 4, and 5 are proposed in the following. In this chapter, the inner product between two vectors a and b is denoted by $\langle a, b \rangle$.

Lemma 2.

$$
\langle \boldsymbol{W}(t), \widetilde{\boldsymbol{U}}(t) \rangle = 0, \ \forall t.
$$

The proof of Lemma 2 is provided in Appendix D.2.

Lemma 3. *Under the GB-PANDAS routing policy, for any arrival rate vector strictly inside the outer bound of the capacity region, $\boldsymbol{\lambda} \in \Lambda$, and the*

corresponding workload vector of servers $\boldsymbol{w}$ *defined in* (5.8)*, we have the following for any* t_0*:*

$$\mathbb{E}\Big[\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle \Big| Z(t_0)\Big] \leq 0, \ \forall t \geq 0.$$

The proof of Lemma 3 is provided in Appendix D.3.

Lemma 4. *Under the GB-PANDAS routing policy, for any arrival rate vector strictly inside the outer bound of the capacity region,* $\boldsymbol{\lambda} \in \Lambda$*, and the corresponding workload vector of servers* $\boldsymbol{w}$ *defined in* (5.8) *there exists* $T_0 > 0$ *such that for any* $T \geq T_0$ *we have the following:*

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big| Z(t_0)\right]\\ \leq& -\theta_0 T||\boldsymbol{Q}(t_0)||_1 + c_0, \ \forall t_0 \geq 0,\end{aligned}$$

where the constants $\theta_0, c_0 > 0$ *are independent of* $Z(t_0)$*.*

The proof of Lemma 4 is provided in Appendix D.4.

Lemma 5. *Under the GB-PANDAS routing policy, for any arrival rate vector strictly inside the outer bound of the capacity region,* $\boldsymbol{\lambda} \in \Lambda$*, and any* $\theta_1 \in (0,1)$*, there exists* $T_1 > 0$ *such that the following is true for any* $T \geq T_1$ *and for any* $t_0 \geq 0$*:*

$$\begin{aligned}&\mathbb{E}\Big[||\boldsymbol{\Psi}(t_0+T)||_1 - ||\boldsymbol{\Psi}(t_0)||_1\Big| Z(t_0)\Big]\\ \leq& -\theta_1||\boldsymbol{\Psi}(t_0)||_1 + MT,\end{aligned}$$

where $||.||_1$ *is* L^1*-norm.*

The proof of Lemma 5 is provided in Appendix D.6.

We choose the following Lyapunov function, $V : \mathcal{P} \to \mathcal{R}^+$:

$$V(Z(t)) = ||\boldsymbol{W}(t)||^2 + ||\boldsymbol{\Psi}(t)||_1,$$

where $||.||$ and $||.||_1$ are the L^2 and L^1-norm, respectively. Then,

$$
\begin{aligned}
&\mathbb{E}\Big[V(Z(t_0+T)) - V(Z(t_0))\Big|Z(t_0)\Big] \\
&= \mathbb{E}\Big[||\boldsymbol{W}(t_0+T)||^2 - ||\boldsymbol{W}(t_0)||^2\Big|Z(t_0)\Big] \\
&\quad + \mathbb{E}\Big[||\boldsymbol{\Psi}(t_0+T)||_1 - ||\boldsymbol{\Psi}(t_0)||_1\Big|Z(t_0)\Big] \\
&\overset{(a)}{=} \mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(||\boldsymbol{W}(t+1)||^2 - ||\boldsymbol{W}(t)||^2\Big)\Big|Z(t_0)\right] \\
&\quad + \mathbb{E}\Big[||\boldsymbol{\Psi}(t_0+T)||_1 - ||\boldsymbol{\Psi}(t)||_1\Big|Z(t_0)\Big] \\
&\overset{(b)}{=} \mathbb{E}\Bigg[\sum_{t=t_0}^{t_0+T-1}\Big(||\boldsymbol{A}(t) - \boldsymbol{S}(t) + \widetilde{\boldsymbol{U}}(t)||^2 \\
&\qquad\qquad + 2\langle \boldsymbol{W}(t), \boldsymbol{A}(t) - \boldsymbol{S}(t)\rangle + 2\langle \boldsymbol{W}(t), \widetilde{\boldsymbol{U}}(t)\rangle\Big)\Big|Z(t_0)\Bigg] \\
&\quad + \mathbb{E}\Big[||\boldsymbol{\Psi}(t_0+T)||_1 - ||\boldsymbol{\Psi}(t)||_1\Big|Z(t_0)\Big] \\
&\overset{(c)}{\leq} 2\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t) - \boldsymbol{S}(t)\rangle\Big)\Big|Z(t_0)\right] \\
&\quad + \mathbb{E}\Big[||\boldsymbol{\Psi}(t_0+T)||_1 - ||\boldsymbol{\Psi}(t)||_1\Big|Z(t_0)\Big] + c_1 \\
&\overset{(d)}{=} 2\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|Z(t_0)\right] \\
&\quad + 2\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|Z(t_0)\right] \\
&\quad + \mathbb{E}\Big[||\boldsymbol{\Psi}(t_0+T)||_1 - ||\boldsymbol{\Psi}(t)||_1\Big|Z(t_0)\Big] + c_1,
\end{aligned}
\tag{5.11}
$$

where (a) is true by the telescoping property, (b) follows by the dynamic of $\boldsymbol{W}(.)$ derived in (5.10), (c) follows by Lemma 2 and the fact that the task arrival is assumed to be bounded and the service and unused service are also bounded as the number of servers are finite, so the pseudo arrival, service, and unused service are also bounded, and therefore there exists a constant c_1 such that $||\boldsymbol{A}(t) - \boldsymbol{S}(t) + \widetilde{\boldsymbol{U}}(t)||^2 \leq \frac{c_1}{T}$, and (d) follows by adding and subtracting the intermediary term $\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle$.

By choosing $T \geq \max\{T_0, T_1, \frac{\theta_1}{2\theta_0}\}$ and using Lemmas 3, 4, and 5, the drift

of the Lyapunov function in (5.11) is the following:

$$\begin{aligned}&\mathbb{E}\Big[V(Z(t_0+T))-V(Z(t_0))\Big|Z(t_0)\Big]\\&\leq -\theta_1\Big(||\boldsymbol{Q}(t_0)||_1+||\boldsymbol{\Psi}(t_0)||_1\Big)+c_2,\ \forall t_0,\end{aligned}$$

where $c_2 = 2c_0 + c_1 + MT$.
By choosing any positive constant $\theta_2 > 0$ let $\mathcal{P} = \big\{Z = \big(\boldsymbol{Q}, \boldsymbol{\eta}, \boldsymbol{\Psi}\big) \in \mathcal{S} : ||\boldsymbol{Q}||_1 + ||\boldsymbol{\Psi}||_1 \leq \frac{\theta_2+c}{\theta_1}\big\}$, where $\mathcal{P}$ is a finite set of the state space. By this choice of $\mathcal{P}$, the condition (5.6) in the extended version of the Foster-Lyapunov theorem holds by choices of $\theta = \theta_1$ and $C = c_2$, so the positive recurrence proof of the Markov chain and the throughput optimality proof of the GB-PANDAS algorithm are completed. Note that a corollary of this result is that Λ is the capacity region of the system. □

Note that in the proof of throughput optimality, we do not rely on the fact of using prioritized scheduling. Therefore, for the purpose of throughput optimality, an idle server can serve any task in its N sub-queues as 1-local, 2-local, $\cdots$, and N-local tasks decrease the expected workload at the same rate. The prioritized scheduling is to minimize the mean task completion time experienced by tasks, which will be of interest in heavy-traffic optimality. If fairness among jobs is of interest, we can assume sub-queues associated to jobs in each server and schedule an idle server to serve a task of the job which has the highest priority in terms of fairness. This does not affect the stability of the system.

5.4 The Affinity System Model

Consider M unit-rate multi-skilled servers and N_T number of task types as depicted in Figure 5.3. The set of servers and task types are denoted by $\mathcal{M} = \{1, 2, \cdots, M\}$ and $\mathcal{L} = \{1, 2, \cdots, N_T\}$, respectively. Each task can be processed by any of the M servers, but with possibly different rates. The service times are assumed to be non-preemptive and discrete valued with an unknown distribution. Non-preemptive service means that the central load balancing algorithm cannot interrupt an in-service task, i.e. no other task is scheduled to a server until the server completely processes the task that is currently receiving service. The extension of the analysis for continuous

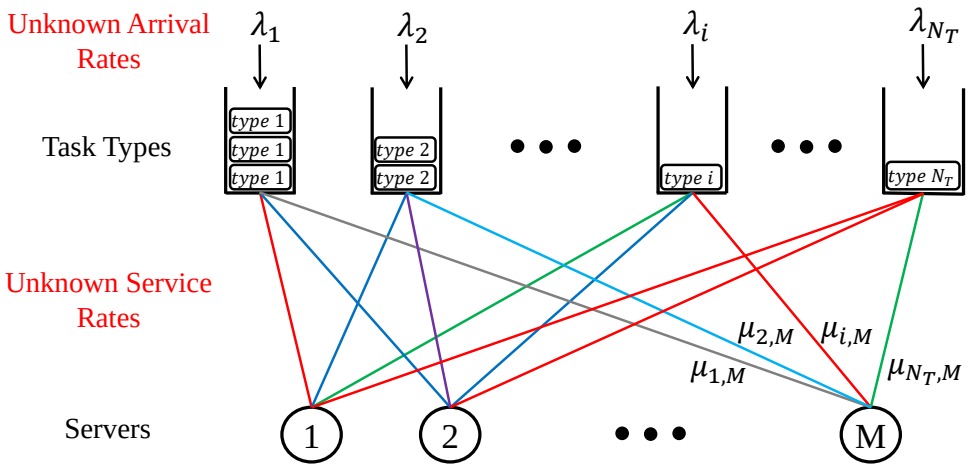

Figure 5.3: Affinity scheduling setup with multi-type tasks and multi-skilled servers.

service time, using approximation methods of continuous distributions with discrete ones, is an interesting future work. In this discrete time model, time is indexed by $t \in \mathbb{N}$. In the following, service time distributions and task arrivals are discussed, which are both unknown to the central scheduler.

Service time distribution: The service time offered by server $m \in \mathcal{M}$ to task type $i \in \mathcal{L}$ is a discrete-type random variable with cumulative distribution function (CDF) $F_{i,m}$ with mean $\frac{1}{\mu_{i,m}}$ or correspondingly with rate $\mu_{i,m} > 0$. The service time distribution does not require the memoryless property. We further assume that the support of the service time is bounded, which is a realistic assumption and reduces the unnecessary complexity of the proofs specially in Lemma 9. The extension of the analysis for service times with unbounded supports is an interesting future work. Note that the completion time for a task is the waiting time for that task until it is scheduled to a server plus the service time of the task on the server. Waiting time depends on the servers' status, the queue lengths or more specifically other tasks that are in the system or may arrive later, and the load balancing algorithm that is used, while service time has the mentioned distribution.

Task arrival: The number of incoming tasks of type $i \in \mathcal{L}$ at the beginning of time slot t is a random variable on non-negative integer numbers that is denoted by $A_i(t)$, which are temporarily identically distributed and independent from each other. Denote the arrival rate of task type i by λ_i, i.e. $\mathbb{E}[A_i(t)] = \lambda_i$. In the stability proof of Blind GB-PANDAS we need λ_i to be strictly positive, so without loss of generality we exclude task types

with zero arrival rate from $\mathcal{L}$. Furthermore, we assume that the number of each incoming task type at a time slot is bounded by constant C_A and is zero with positive probability, i.e. $P(A_i(t) < C_A) = 1$ and $P(A_i(t) = 0) > 0$ for any $i \in \mathcal{L}$. The set of arrival rates for all task types is denoted by vector $\boldsymbol{\lambda} = (\lambda_i : i \in \mathcal{L})$.

Affinity scheduling problem refers to load balancing for such a system described above. The fluid model planning algorithm [121], MaxWeight [127], and cμ-rule [124] are the baseline algorithms for affinity scheduling. All these algorithms in addition to GB-PANDAS use the rate of service times instead of the CDF functions. Hence, the system model can be summarized as an $N_T \times M$ matrix, where element (i, m) is the processing rate of task type i on server m, $\mu_{i,m}$, as follows:

$$B_\mu = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \mu_{1,3} & \cdots & \mu_{1,M} \\ \mu_{2,1} & \mu_{2,2} & \mu_{2,3} & \cdots & \mu_{2,M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_{N_T,1} & \mu_{N_T,2} & \mu_{N_T,3} & \cdots & \mu_{N_T,M} \end{bmatrix}_{N_T,M} . \tag{5.12}$$

If both the set of arrival rates $\boldsymbol{\lambda} = (\lambda_i : i \in \mathcal{L})$ and the service rate matrix B_μ are known, the fluid model planning algorithm [121] derives the delay optimal load balancing by solving a linear programming. However, if the arrival rates of task types are not known, the delay optimal algorithm becomes an open problem which has not been solved for more than three decades. MaxWeight [127] and cμ-rule [124] can be used for different objectives when we do not know the arrival rates, but none have delay optimality. In this work, we are assuming that we lack knowledge of not only the arrival rates $\boldsymbol{\lambda}$, but also the service rate matrix B_μ. We take an exploration and exploitation approach to make our estimation of the underlying model, which is the service rate matrix, more accurate, and to keep the system stable.

5.4.1 The Queueing Structure of the GB-PANDAS Algorithm in the Affinity Problem

Every algorithm has its own specific queueing structure. For example, there is only a single central queue for the First-Come-First-Served (FCFS) algorithm, but there are N_T number of queues when using fluid model planning,

Max-Weight, or cμ-rule. In the following, we present the queueing structure used for GB-PANDAS that captures the trade-off between the complexity of the system model and the complexity of the queueing structure very well. What we mean by the complexity of the system model is the heterogeneity of the service rate matrix, e.g. if all the elements of this matrix are the same number, the system is less complex than the case where each element of the matrix is different from other elements of the matrix.

The heterogeneity of the system from the perspective of server m is captured in the m^{th} column of the service rate matrix. Consider the m^{th} column of the matrix has N^m distinct values, where N^m can be any number from 1 to N_T. It is obvious that any of the task types with the same service (processing) rate on server m look the same from the perspective of this server. Denote the N^m distinct values of the m^{th} column of B_μ by $\{\alpha_m^1, \alpha_m^2, \cdots, \alpha_m^{N^m}\}$ and without loss of generality assume that $\alpha_m^1 > \alpha_m^2 > \cdots > \alpha_m^{N^m}$. We call all the task types with a processing rate of α_m^n on the m^{th} server, the n-local tasks to that server, and denote them by $\mathcal{L}_m^n = \{i \in \mathcal{L} : \mu_{i,m} = \alpha_m^n\}$. For ease of notation, we use both $\mu_{i,m}$ and α_m^n throughout the chapter interchangeably; however, they are in fact capturing the same phenomenon, but with different interpretations. Note that the n-local tasks to server m can be called (n, m)-local tasks in order to place more emphasis on the pair n and m, so the n-local tasks to server m are not necessarily the same as the n-local tasks to server m'. We allocate N^m queues for server m, where the n^{th} queue of server m holds all task types that are routed to this server and are n-local to it. As depicted in Figure 5.4, different servers can have different numbers of queues since the heterogeneity of the system model can be different from the perspective of different servers. We may interchangeably use queue or sub-queue to refer to the n^{th} queue (sub-queue) of the m^{th} server. The N^m sub-queues of the m^{th} server are denoted by $Q_m^1, Q_m^2, \cdots, Q_m^{N^m}$ and the queue lengths of these sub-queues, defined as the number of tasks in these sub-queues, at time slot t are denoted by $Q_m^1(t), Q_m^2(t), \cdots, Q_m^{N^m}(t)$.

In the next subsection, the GB-PANDAS algorithm is proposed when the service rate matrix B_μ is known. Balanced-PANDAS for a data center with three levels of data locality is proposed by [134], and here we are proposing the Generalized Balanced-PANDAS algorithm from another perspective which is of its own interest.

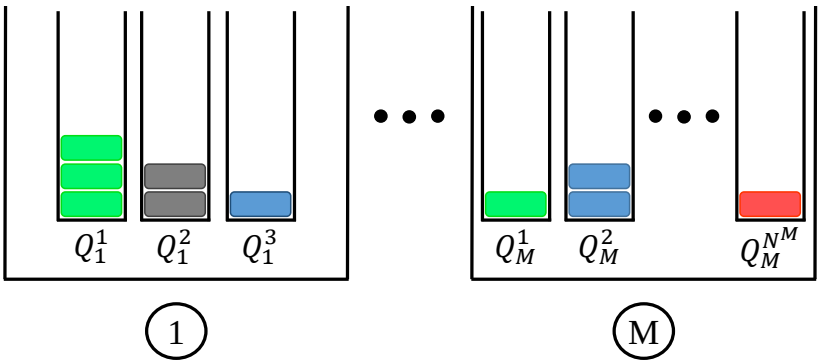

Figure 5.4: The queueing structure for the GB-PANDAS algorithm.

5.4.2 The GB-PANDAS Algorithm with Known Service Rate Matrix B_μ for the Affinity Problem

Before getting into the GB-PANDAS algorithm, we need to define the workload on server m.

Definition 19. *The average time needed for server m to process all tasks queued in its N^m sub-queues at time slot t is defined as the workload on the server:*

$$W_m(t) = \frac{Q_m^1(t)}{\alpha_m^1} + \frac{Q_m^2(t)}{\alpha_m^2} + \cdots + \frac{Q_m^{N^m}(t)}{\alpha_m^{N^m}}. \tag{5.13}$$

A load balancing algorithm consists of two parts, routing and scheduling. The routing policy determines the queue at which an incoming task is stored until it is assigned to a server for service. When a server becomes idle, the scheduling policy determines the next task that receives service on the idle server. The routing and scheduling policies of the GB-PANDAS algorithm are as follows:

GB-PANDAS Routing Policy: An incoming task of type i is routed to the corresponding sub-queue of the server with the minimum weighted workload, where ties are broken arbitrarily to the favor of the fastest server. The server m^* with the minimum weighted workload is defined as

$$m^* = \arg\min_{m \in \mathcal{M}} \frac{W_m(t)}{\mu_{i,m}}.$$

The corresponding sub-queue of server m^* for a task of type i is n if $\mu_{i,m} = \alpha_m^n$.

GB-PANDAS Scheduling Policy: An idle server m at time slot t is scheduled to process a task of sub-queue Q_m^1 if there is any. If $Q_m^1(t) = 0$, a

task of sub-queue Q_m^2 is scheduled to the server, and so on. It is a common assumption that servers do not have the option of processing the tasks queued in front of other servers, so a server remains idle if all its sub-queues are empty. Note that the routing policy is doing a sort of weighted water-filling for workloads, so the probability that a server becomes idle goes to zero as the load increases at heavy traffic regime. Remember that the tasks in sub-queue Q_m^1 are the fastest types of tasks for server m, the tasks in sub-queue Q_m^2 are the second fastest, and so on. Using this priority scheduling, the faster tasks in the N^m sub-queues of server m are processed first. Given the minimum weighted workload routing policy, the priority scheduling is optimal as it minimizes the mean task completion time of all tasks in the N^m sub-queues of server m. In the following, Max-Weight and cμ-rule algorithms are discussed for the sake of completeness.

Remark 10. *Prioritized scheduling has no effect in the throughput-optimality proof of the GB-PANDAS algorithm and a work-conserving scheduling of a server to its sub-queues suffices for the purpose of system's stability. As a result, the GB-PANDAS policy can be implemented by considering a single queue per server at the expense of losing priority scheduling. In a single queue per server structure, instead of maintaining a server's sub-queue lengths, the workload of the server defined in* (5.13) *is maintained. At the arrival of an n-local task to server m, the server's workload is increased by $\frac{1}{\alpha_m^n}$, instead of increasing the corresponding sub-queue's length by one, and the workload is decreased at the departure of a task by its corresponding load.*

5.4.3 The Max-Weight and c-μ-Rule Algorithms with Known Service Rate Matrix B_μ

The queueing structure used for Max-Weight and c-μ-rule is as depicted in Figure 5.3, where there is a separate queue for each type of task. Denote the N_T queues by $Q_1, Q_2, \cdots, Q_{N_T}$, and their corresponding queue lengths at time slot t by $Q_1(t), Q_2(t), \cdots, Q_{N_T}(t)$. Note that the GB-PANDAS algorithm requires $M \times N_T$ number of queues in the worst case scenario, but it can use the symmetry of specific real-world structures to decrease the number of queues dramatically. As an example, for servers with rack structures, where Hadoop is used for MapReduce data placement with three replicas

of data chunks on severs, Max-Weight and c-μ-rule require $\binom{M}{3} = O(M^3)$ number of queues, while GB-PANDAS requires $3M$ queues. A task is routed to a server at the time of its arrival under the GB-PANDAS algorithm, while a task waits in its queue under both Max-Weight and c-μ-rule algorithms, waiting to be scheduled for service, which is discussed below.

Max-Weight Scheduling Policy: An idle server m at time slot t is scheduled to process a task of type j from Q_j, if there is any, such that

$$j \in \arg\max_{i \in \mathcal{L}} \{\mu_{i,m} \cdot Q_i(t)\}.$$

The Max-Weight algorithm is throughput-optimal, but it is not heavy-traffic or delay optimal.

C-μ-rule Scheduling Policy: Consider that queue Q_i incurs a cost of $C_i\big(Q_i(t)\big)$ at time slot t, where $C_i(.)$ is increasing and strictly convex. The c-μ-rule algorithm maximizes the rate of decrease of the instantaneous cost at all time slots by the following scheduling policy. An idle server m at time slot t is scheduled to process a task of type j from Q_j, if there is any, such that

$$j \in \arg\max_{i \in \mathcal{L}} \big\{\mu_{i,m} \cdot C_i'\big(Q_i(t)\big)\big\},$$

where $C'(.)$ is the first derivative of the cost function. The c-μ-rule algorithm minimizes both instantaneous and cumulative queueing costs, asymptotically. The mean task completion time corresponds to linear cost functions for all task types, so c-μ-rule cannot minimize the mean task completion time, and as the result, is not heavy-traffic optimal.

5.4.4 Capacity Region of the Affinity Problem

We propose a decomposition of the arrival rate vector $\boldsymbol{\lambda} = (\lambda_i : i \in \mathcal{L})$ as follows. For any task type $i \in \mathcal{L}, \lambda_i$ is decomposed into $(\lambda_{i,m}, m \in \mathcal{M})$, where $\lambda_{i,m}$ is assumed to be the arrival rate of type i tasks for server m. Hence, $\lambda_i = \sum_{m=1}^{M} \lambda_{i,m}$. By using the fluid model planning algorithm, the affinity queueing system can be stabilized under a given arrival rate vector $\boldsymbol{\lambda}$ as long as the necessary condition of total 1-local, 2-local, ..., and N^m local

load on server m being strictly less than one for any server m is satisfied:

$$\sum_{i\in\mathcal{L}} \frac{\lambda_{i,m}}{\mu_{i,m}} < 1, \qquad \forall m \in \mathcal{M}. \tag{5.14}$$

Hence, the capacity region of the affinity problem is the set of all arrival rate vectors $\boldsymbol{\lambda}$ that has a decomposition $(\lambda_{i,m}, i \in \mathcal{L}, m \in \mathcal{M})$ satisfying (5.14):

$$\begin{aligned} \Lambda = \big\{\boldsymbol{\lambda} = (\lambda_i : i \in \mathcal{L}) \;\big|\; \exists \lambda_{i,m} \geq 0, \forall i \in \mathcal{L}, \forall m \in \mathcal{M}, s.t. \\ \lambda_i = \sum_{m=1}^{M} \lambda_{i,m},\ \forall i \in \mathcal{L}, \quad \sum_{i\in\mathcal{L}} \frac{\lambda_{i,m}}{\mu_{i,m}} < 1, \forall m \in \mathcal{M}\big\}. \end{aligned} \tag{5.15}$$

A linear programming optimization problem can be solved to find the capacity region Λ. The GB-PANDAS algorithm stabilizes the system for any arrival rate vector inside the capacity region by knowing the service rate matrix. It is proved in Section 5.6 that the Blind GB-PANDAS algorithm is throughput-optimal without the knowledge of the service rate matrix, B_μ.

5.5 The Blind GB-PANDAS Algorithm for the Affinity Problem

The GB-PANDAS and Max-Weight algorithms need to know the precise value of the service rate matrix, but this requirement is not realistic for real applications. Furthermore, the service rate matrix can change over time, which confuses the load balancing algorithm if it uses a fixed given service rate matrix. In the Blind version of GB-PANDAS, the service rate matrix is initiated randomly and is updated as the system is running. More specifically, an exploration-exploitation framework is combined with GB-PANDAS. In the exploration phase, the routing and scheduling are performed so as to allow room for making the estimations of the system parameters more precise, and in the exploitation phase the routing and scheduling are done based on the available estimation of the service rate matrix so as to stabilize the system. Here we assume that N^m is known as well as the locality level of a task on servers that can be inferred from prior knowledge on the structure of the system. This is not a necessary assumption for throughput-optimality

proof, but it makes the intuition behind Blind GB-PANDAS more clear. As mentioned before, a single queue per server can be used when using Blind GB-PANDAS, in which case, there is no need to know N^m as well as the ordering of service rates offered by servers for different task types.

We first propose the updating method used for the service rate matrix before getting into the routing and scheduling policies of the Blind GB-PANDAS algorithm. The estimated service rate matrix at time slot t is denoted as

$$\widetilde{B}_\mu(t) = \begin{bmatrix} \widetilde{\mu}_{1,1}(t) & \widetilde{\mu}_{1,2}(t) & \widetilde{\mu}_{1,3}(t) & \dots & \widetilde{\mu}_{1,M}(t) \\ \widetilde{\mu}_{2,1}(t) & \widetilde{\mu}_{2,2}(t) & \widetilde{\mu}_{2,3}(t) & \dots & \widetilde{\mu}_{2,M}(t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \widetilde{\mu}_{N_T,1}(t) & \widetilde{\mu}_{N_T,2}(t) & \widetilde{\mu}_{N_T,3}(t) & \dots & \widetilde{\mu}_{N_T,M}(t) \end{bmatrix}. \tag{5.16}$$

Note that $\widetilde{\alpha}_m^1(t), \widetilde{\alpha}_m^2(t), \cdots, \widetilde{\alpha}_m^{N^m}(t),\ \forall m \in \mathcal{M}$, which are the estimates of $\alpha_m^1(t), \alpha_m^2(t), \cdots, \alpha_m^{N^m}(t),\ \forall m \in \mathcal{M}$ at time slot t, are nothing but the distinct values of the elements of the service rate matrix. More specifically, those are the $\widetilde{\alpha}_m^n,\ \forall m \in \mathcal{M},\ \forall n \in \{1, 2, \cdots, N^m\}$ that are getting updated and then mapped into their corresponding elements in the service rate matrix to form $\widetilde{B}_\mu$ in (5.16) as mentioned in Subsection 5.4.1. Consider a random initialization of $\widetilde{\alpha}_m^n(0) > 0,\ \forall m \in \mathcal{M},\ \forall n \in \{1, 2, \cdots, N^m\}$ at time slot $t = 0$. If server m has processed $\widetilde{n} - 1$ tasks that are n-local to this server by time t_1, the estimate of α_m^n at this time slot is $\widetilde{\alpha}_m^n(t_1)$, and a new observation of service time for n-local task to server m is made at time slot $t_2 > t_1$ as $T_m^n(t_2)$, we have $\widetilde{\alpha}_m^n(t) = \widetilde{\alpha}_m^n(t_1)$ for $t_1 \leq t < t_2$ and the update of this parameter at time slot t_2 is

$$\widetilde{\alpha}_m^n(t_2) = \frac{\widetilde{n} - 1}{\widetilde{n}} \cdot \widetilde{\alpha}_m^n(t_1) + \frac{1}{\widetilde{n} \cdot T_m^n(t_2)}. \tag{5.17}$$

Note that $\widetilde{\alpha}_m^n$ is the service rate, not the service time mean, that is why $\frac{1}{T_m^n(t_2)}$ is used above in the update of the service rate. In the following, the routing and scheduling policies of Blind GB-PANDAS are presented, where the exploration rate is chosen in such a way that infinitely many n-local tasks are scheduled for service on server m for any $m \in \mathcal{M}$ and any $n \in \{1, 2, \cdots, N^m\}$ so that by using the strong law of large numbers, the parameter estimations in (5.17) converge to their real values almost surely.

Blind GB-PANDAS Routing Policy: The estimated workload on server m at time slot t is defined based on parameter estimations in (5.17) as

$$\widetilde{W}_m(t) = \frac{Q_m^1(t)}{\widetilde{\alpha}_m^1(t)} + \frac{Q_m^2(t)}{\widetilde{\alpha}_m^2(t)} + \cdots + \frac{Q_m^{N^m}(t)}{\widetilde{\alpha}_m^{N^m}(t)}. \tag{5.18}$$

The routing of an incoming task is based on the following exploitation policy with probability $p_e = max(1 - p(t), 0)$, and is based on the exploration policy otherwise, where $p(t) \to 0$ as $t \to \infty$ and $\sum_{t=0}^{\infty} p(t) = \infty$, e.g. the exploitation probability can be chosen as $p_e = 1 - \frac{1}{t^c}$ for $0 < c \leq 1$.

- **Exploitation phase:** An incoming task of type i is routed to the corresponding sub-queue of the server with the minimum estimated weighted workload, where ties are broken arbitrarily. The server $\widetilde{m}^*$ with the minimum weighted workload for task of type i is defined as

  $$\widetilde{m}^* = \arg\min_{m \in \mathcal{M}} \frac{\widetilde{W}_m(t)}{\widetilde{\mu}_{i,m}(t)}.$$

  The corresponding sub-queue of server $\widetilde{m}^*$ for a task of type i is n if $\widetilde{\mu}_{i,\widetilde{m}^*} = \widetilde{\alpha}_{\widetilde{m}^*}^n$.

- **Exploration phase:** An incoming task of type i is routed to the corresponding sub-queue of a server chosen uniformly at random among $\{1, 2, \cdots, M\}$.

Blind GB-PANDAS Scheduling Policy: The scheduling of an idle server is based on the following exploitation policy with probability p_e, and is based on the exploration policy otherwise.

- **Exploitation phase:** Priority scheduling is performed for an idle server as discussed in Subsection 5.4.2. We emphasize that given the routing policy, priority scheduling is the optimal scheduling policy in terms of minimizing the average completion time of tasks.

- **Exploration phase:** An idle server is scheduled to one of its non-empty sub-queues uniformly at random, and stays idle if all its sub-queues are empty.

Since the arrival rate of any task type is strictly positive, infinitely many of each task type arrives to system, and given the fact that the probability of exploration in both routing and scheduling policies decays such that

$\sum_{t=0}^{\infty} p(t) = \infty$, using the second Borel-Cantelli lemma (zero-one law), it is obvious that n-local tasks to server m are scheduled to this server for infinitely many times for any locality level and any server, so $\widetilde{B}_\mu(t) \to B_\mu$ as $t \to \infty$ using the updates in (5.17).

Remark 11. *There has been a debate in the queueing community whether the exploration phase in a load balancing algorithm is required to stabilize a queueing system with unknown processing rates or the processing rates are learned through a natural learning phenomenon and, as a result, no exploration is needed. We provide an example in Figure 5.5 that shows no exploration can not only increase the mean task completion time, but it can also make the system unstable when the arrival rates are inside the capacity region of the queueing system. Consider a queueing system as depicted on the left-hand side of Figure 5.5, where the processing times of any tasks on any servers are deterministic with the given rates and the arrival process of tasks is deterministic as well with the rates shown in the figure. It is obvious that the optimal load balancing is to process task type 1 on server 1 and task type 2 on server 2. However, if the processing rates are initialized as in the middle queueing system of Figure 5.5, for any* $\lambda_1 \leq 0.5$ *and* $\lambda_2 \leq 0.5$*, task type 1 is processed by server 2 and task type 2 is processed by server 1 under the GB-PANDAS and MaxWeight algorithms, resulting in a mean task completion time that is twice the optimal value. On the other hand, if the processing rates are initialized as in the right-hand-side queueing system of Figure 5.5, for any* $0.5 < \lambda_1 \leq 1$ *and* $0.5 < \lambda_2 \leq 1$*, the system is unstable under the GB-PANDAS and MaxWeight algorithms, while such processing rates are inside the capacity region of the queueing system. As a result, exploration is required in the load balancing algorithm in general for a queueing system with unknown processing rates. Using the intuition of the given example, it is a promising future work to find conditions for which exploration is not required for the purpose of delay optimality and/or stability.*

5.5.1 Queueing Dynamics under the Blind GB-PANDAS Algorithm for the Affinity Problem

Denote the queue length vector at time slot t by $\boldsymbol{Q}(t) = \big(Q_1^1(t), Q_1^2(t), \cdots, Q_1^{N^1}(t), \cdots, Q_M^{N^M}(t)\big)$. Let the number of incoming tasks of type i that are

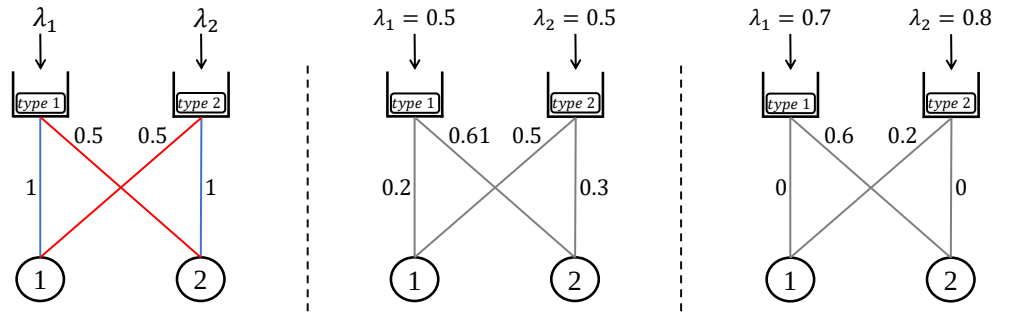

Figure 5.5: This example shows that a queueing system with unknown processing rates can even be unstable for some initialization of processing rates if there is no exploration in the load balancing algorithm.

routed to their corresponding sub-queue of server m at the beginning of time slot t be denoted as $A_{i,m}(t)$. Then, by denoting the number of incoming n-local tasks to server m that are routed to Q_m^n at the beginning of time slot t by $A_m^n(t)$, we have:

$$A_m^n(t) = \sum_{i \in \mathcal{L}_m^n} A_{i,m}(t), \ \forall m \in \mathcal{M}, \ 1 \leq n \leq N^m. \tag{5.19}$$

Denote the set of working status of servers by vector $\boldsymbol{f}(t) = \big(f_1(t), f_2(t), \cdots, f_M(t)\big)$, where

$$f_m(t) = \begin{cases} -1, \text{ if server } m \text{ is idle,} \\ 1, \text{ if server } m \text{ processes a 1-local task from } Q_m^1, \\ 2, \text{ if server } m \text{ processes a 2-local task from } Q_m^2, \\ \vdots \\ N^m, \text{ if server } m \text{ processes an } N^m\text{-local task from } Q_m^{N^m}. \end{cases}$$

If server m finishes processing a task at the end of time slot $t-1$, i.e. $f_m(t^-) = -1$, a scheduling decision is taken for time t based on $\boldsymbol{Q}(t)$ and $\boldsymbol{f}(t)$. Denote the scheduling decision for server m at time slot t by $\eta_m(t)$ that is defined as follows. For all busy servers, $\eta_m(t) = f_m(t)$, and when $f_m(t^-) = -1$, i.e. server m is idle, $\eta_m(t)$ is determined by the scheduler according to the Blind GB-PANDAS algorithm. Let $\boldsymbol{\eta}(t) = \big(\eta_1(t), \eta_2(t), \cdots, \eta_M(t)\big)$.

Let $S_m^n(t)$ denote the n-local service provided by server m, where such a service has the rate of α_m^n if $\eta_m(t) = n$ for $1 \leq n \leq N^m$, and the rate is zero otherwise. Then, the queue dynamics for any $m \in \mathcal{M}$ is as follows:

$$\begin{aligned} &Q_m^n(t+1) = Q_m^n(t) + A_m^n(t) - S_m^n(t), \text{ for } 1 \leq n \leq N^m - 1, \\ &Q_m^{N^m}(t+1) = Q_m^{N^m}(t) + A_m^{N^m}(t) - S_m^{N^m}(t) + U_m(t), \end{aligned} \tag{5.20}$$

where $U_m(t) = \max\left\{0, S_m^{N^m}(t) - A_m^{N^m}(t) - Q_m^{N^m}(t)\right\}$ is the unused service offered by server m at time slot t.

Note that $\left\{\boldsymbol{Q}(t),\, t \geq 0\right\}$ does not necessarily form a Markov chain, i.e. $\boldsymbol{Q}(t+1)|\boldsymbol{Q}(t) \not\perp \boldsymbol{Q}(t-1)$, since nothing can be said about locality of an in-service task at a server by just knowing the queue lengths. Even $\left\{\left(\boldsymbol{Q}(t), \boldsymbol{\eta}(t)\right),\, t \geq 0\right\}$ is not a Markov chain since the service time distributions do not necessarily have the memoryless property. In order to use Foster-Lyapunov theorem for proving the positive recurrence of a Markov chain, we need to consider another measurement of the status of servers as follows.

- Let $\Psi_m(t)$ denote the number of time slots at the beginning of time slot t that server m has been allocated on the current in-service task on server m. This parameter is set to zero when server m finishes processing a task. Let $\boldsymbol{\Psi}(t) = \big(\Psi_1(t), \Psi_2(t), \cdots, \Psi_M(t)\big)$.

Lemma 6. *$\left\{\boldsymbol{Z}(t) = \big(\boldsymbol{Q}(t), \boldsymbol{\eta}(t), \boldsymbol{\Psi}(t)\big), t \geq 0\right\}$ forms an irreducible and aperiodic Markov chain. The state space of the Markov chain $\{\boldsymbol{Z}(t)\}$ is $\mathcal{S} = \big(\prod_{m\in\mathcal{M}} \mathbb{N}^{N^m}\big) \times \big(\prod_{m\in\mathcal{M}} \{1, 2, \cdots, N^m\}\big) \times \mathbb{N}^M$.*

The proof of Lemma 6 is provided in Appendix D.7.

5.6 Throughput Optimality of the Blind GB-PANDAS Algorithm for the Affinity Problem

Subsection 5.6.1 provides preliminaries on the workload dynamic of servers, the ideal workload on servers, some lemmas, and an extended version of the Foster-Lyapunov. The throughput-optimality theorem of the Blind GB-PANDAS algorithm and its proof are presented in Subsection 5.6.2, where

the proof is followed by using Lemmas 7, 8, 9, 10, and 11. Refer to Appendix D for the proofs of all lemmas.

5.6.1 Preliminary Materials and Lemmas

The workload on server m evolves as follows:

$$\begin{aligned}
W_m(t+1) &= \frac{Q_m^1(t+1)}{\alpha_m^1} + \frac{Q_m^2(t+1)}{\alpha_m^2} + \cdots + \frac{Q_m^{N^m}(t+1)}{\alpha_m^{N^m}} \\
&\overset{(a)}{=} \frac{Q_m^1(t) + A_m^1(t) - S_m^1(t)}{\alpha_m^1} + \frac{Q_m^2(t) + A_m^2(t) - S_m^2(t)}{\alpha_m^2} \\
&\quad + \cdots + \frac{Q_m^{N^m}(t) + A_m^{N^m}(t) - S_m^{N^m}(t) + U_m(t)}{\alpha_m^{N^m}} \\
&= W_m(t) + \left(\frac{A_m^1(t)}{\alpha_m^1} + \frac{A_m^2(t)}{\alpha_m^2} + \cdots + \frac{A_m^{N^m}(t)}{\alpha_m^{N^m}}\right) \\
&\quad - \left(\frac{S_m^1(t)}{\alpha_m^1} + \frac{S_m^2(t)}{\alpha_m^2} + \cdots + \frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right) + \frac{U_m(t)}{\alpha_m^{N^m}} \\
&\overset{(b)}{=} W_m(t) + A_m(t) - S_m(t) + \widetilde{U}_m(t),
\end{aligned}$$

where (a) is true by using the queue dynamics in (5.20) and (b) follows from defining the pseudo task arrival, service, and unused services of server m as

$$\begin{aligned}
A_m(t) &= \frac{A_m^1(t)}{\alpha_m^1} + \frac{A_m^2(t)}{\alpha_m^2} + \cdots + \frac{A_m^{N^m}(t)}{\alpha_m^{N^m}}, \ \forall m \in \mathcal{M}, \\
S_m(t) &= \frac{S_m^1(t)}{\alpha_m^1} + \frac{S_m^2(t)}{\alpha_m^2} + \cdots + \frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}, \ \forall m \in \mathcal{M}, \\
\widetilde{U}_m(t) &= \frac{U_m(t)}{\alpha_m^{N^m}}, \ \forall m \in \mathcal{M}.
\end{aligned} \tag{5.21}$$

By defining the pseudo task arrival, service, and unused service processes as $\boldsymbol{A}(t) = \big(A_1(t), A_2(t), \cdots, A_M(t)\big)$, $\boldsymbol{S}(t) = \big(S_1(t), S_2(t), \cdots, S_M(t)\big)$, and $\widetilde{\boldsymbol{U}}(t) = \big(\widetilde{U}_1(t), \widetilde{U}_2(t), \cdots, \widetilde{U}_M(t)\big)$, respectively, the vector of servers' workloads defined by $\boldsymbol{W} = (W_1, W_2, \cdots, W_M)$ evolves as

$$\boldsymbol{W}(t+1) = \boldsymbol{W}(t) + \boldsymbol{A}(t) - \boldsymbol{S}(t) + \widetilde{\boldsymbol{U}}(t). \tag{5.22}$$

Lemma 7. *For any arrival rate vector inside the capacity region, $\boldsymbol{\lambda} \in \Lambda$,*

there exists a load decomposition $\{\lambda_{i,m}\}$ *and* $\delta > 0$ *such that*

$$\sum_{i\in\mathcal{L}} \frac{\lambda_{i,m}}{\mu_{i,m}} < \frac{1}{1+\delta},\ \forall m \in \mathcal{M}. \tag{5.23}$$

The fluid model planning algorithm solves a linear programming to find the load decomposition $\{\lambda_{i,m}\}$ *that is used in its load balancing on the* M *servers. In other words, this load decomposition is a possibility of task assignment on servers to stabilize the system.*

The proof of Lemma 7 is provided in Appendix D.8. Lemma 7 is used in the proof of Lemma 10.

Definition 20. *The ideal workload on server* m *corresponding to the load decomposition* $\{\lambda_{i,m}\}$ *of Lemma 7 is defined as*

$$w_m = \sum_{i\in\mathcal{L}} \frac{\lambda_{i,m}}{\mu_{i,m}},\ \forall m \in \mathcal{M}. \tag{5.24}$$

Let $\boldsymbol{w} = (w_1, w_2, \cdots, w_M)$*. The vector of servers' ideal workload is used as an intermediary term in Lemmas 9 and 10 which are later used for throughput optimality proof of the Blind GB-PANDAS algorithm.*

Lemma 8.

$$\langle \boldsymbol{W}(t), \widetilde{\boldsymbol{U}}(t)\rangle = 0,\ \forall t.$$

The proof of Lemma 8 is provided in Appendix D.9.

The following lemma states that the sum over a time period of the inner product of the workload and the pseudo arrival rate is dominated in an expectation sense by the inner product of the workload and the ideal workload plus constants depending on the initial state of the system.

Lemma 9. *Under the exploration-exploitation routing policy of the Blind GB-PANDAS algorithm, for any arrival rate vector inside the capacity region,* $\boldsymbol{\lambda} \in \Lambda$*, and the corresponding ideal workload vector* $\boldsymbol{w}$ *defined in* (5.24)*, and for any arbitrary small* $\theta_0 > 0$*, there exists* $T_0 > t_0$ *such that for any* $t_0 \geq 0$ *and* $T > T_0$*:*

$$\mathbb{E}\Big[\sum_{t=T_0}^{t_0+T-1} \Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big| \boldsymbol{Z}(t_0)\Big] \leq \theta_0 T \|\boldsymbol{Q}(t_0)\|_1 + c_0,$$

where the constants $\theta_0, c_0 > 0$ are independent of $\boldsymbol{Z}(t_0)$.

The proof of Lemma 9 is provided in Appendix D.10. We emphasize that θ_0 in Lemma 9 can be made arbitrarily small, as can be seen in the proof, which is used in the throughput optimality proof of Blind GB-PANDAS, Theorem 9. Throughout the chapter, $\|.\|$ and $\|.\|_1$ are the L^2-norm and L^1-norm, respectively.

The following lemma is the counterpart of Lemma 9 for the pseudo service process.

Lemma 10. *Under the exploration-exploitation scheduling policy of the Blind GB-PANDAS algorithm, for any arrival rate vector inside the capacity region, $\boldsymbol{\lambda} \in \Lambda$, and the corresponding ideal workload vector $\boldsymbol{w}$ in* (5.24)*, there exists $T_1 > 0$ such that for any $T > T_1$, we have:*

$$\begin{aligned} &\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\ &\leq -\theta_1 T\|\boldsymbol{Q}(t_0)\|_1 + c_1,\ \forall t_0 \geq 0, \end{aligned} \tag{5.25}$$

where the constants $\theta_1, c_1 > 0$ are independent of $\boldsymbol{Z}(t_0)$.

The proof of Lemma 10 is provided in Appendix D.11.

Lemma 11. *Under the exploration-exploitation load balancing of the Blind GB-PANDAS algorithm, for any arrival rate vector inside the capacity region, $\boldsymbol{\lambda} \in \Lambda$, and for any $\theta_2 > 0$, there exists $T_2 > 0$ such that for any $T > T_2$ and for any $t_0 \geq 0$, we have:*

$$\mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0+T)\|_1 - \|\boldsymbol{\Psi}(t_0)\|_1\Big|\boldsymbol{Z}(t_0)\Big] \leq -\theta_2\|\boldsymbol{\Psi}(t_0)\|_1 + MT.$$

The proof of Lemma 11 is provided in Appendix D.12.

Theorem 3.3.8 in [203], an extended version of the Foster-Lyapunov theorem: Consider an irreducible Markov chain $\{\boldsymbol{Z}(t)\}$, where $t \in \mathbb{N}$, with a state space $\mathcal{S}$. If there exists a function $V : \mathcal{S} \to \mathcal{R}^+$, a positive integer $T \geq 1$, and a finite set $\mathcal{P} \subseteq \mathcal{S}$ satisfying the following condition:

$$\begin{aligned} &\mathbb{E}\big[V(Z(t_0+T)) - V(Z(t_0))\big|\boldsymbol{Z}(t_0) = z\big] \\ &\leq -\theta \cdot \mathbb{I}_{\{z\in\mathcal{P}^c\}} + C \cdot \mathbb{I}_{\{z\in\mathcal{P}\}}, \end{aligned} \tag{5.26}$$

for some $\theta > 0$ and $C < \infty$, then the irreducible Markov chain $\{\boldsymbol{Z}(t)\}$ is positive recurrent.

5.6.2 Throughput Optimality Theorem and Proof

Theorem 9. *The Blind GB-PANDAS algorithm is throughput-optimal for a system with affinity setup discussed in Section 5.4, with general service time distributions with finite means and supports, without prior knowledge on the service rate matrix B_μ and the arrival rate vector $\boldsymbol{\lambda}$.*

Proof. We use the Foster-Lyapunov theorem for proving that the irreducible and aperiodic Markov chain $\big\{\boldsymbol{Z}(t) = \big(\boldsymbol{Q}(t), \boldsymbol{\eta}(t), \boldsymbol{\Psi}(t)\big),\ t \geq 0\big\}$ (Lemma 6) is positive recurrent under the Blind GB-PANDAS algorithm, as far as the arrival rate vector is inside the capacity region, $\boldsymbol{\lambda} \in \Lambda$. This means that as time goes to infinity, the distribution of $\boldsymbol{Z}(t)$ converges to its stationary distribution, which implies that the system is stable and Blind GB-PANDAS is throughput-optimal. To this end, we choose the following Lyapunov function $V : \mathcal{S} \to \mathcal{R}^+$ and use Lemmas 7, 8, 9, 10, and 11 to derive its drift afterward:

$$V(\boldsymbol{Z}(t)) = \|\boldsymbol{W}(t)\|^2 + \|\boldsymbol{\Psi}(t)\|_1. \tag{5.27}$$

By choosing θ_0 in Lemma 9 less than θ_1 in Lemma 10, $\theta_0 < \theta_1$, we get T_0 from Lemma 9, which is used in the drift of the Lyapunov function in Lemma 12.

Lemma 12. *For any $t_0 \leq T_0 < T$, specifically T_0 from Lemma 9 that is dictated by choosing $\theta_0 < \theta_1$, we have the following for the drift of the Lyapunov function in* (5.27)*, where T_0 is used in the first summation after the inequality:*

$$\begin{aligned}
&\mathbb{E}\Big[V(\boldsymbol{Z}(t_0+T)) - V(\boldsymbol{Z}(t_0))\Big|\boldsymbol{Z}(t_0)\Big] \\
&\leq 2\mathbb{E}\left[\sum_{t=T_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\quad + 2\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\quad + \mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0+T)\|_1 - \|\boldsymbol{\Psi}(t)\|_1\Big|\boldsymbol{Z}(t_0)\Big] + c_2\|\boldsymbol{Q}(t_0)\|_1 + c_3.
\end{aligned} \tag{5.28}$$

The proof of Lemma 12 is provided in Appendix D.13. By choosing $T > \max\{T_0, T_1, T_2, \frac{\theta_2+c_2}{2(\theta_1-\theta_0)}\}$, where $\theta_2 > 0$ is the one in Lemma 11, and substituting the terms on the right-hand side of the Lyapunov function drift (D.35) in Lemma 12 from the corresponding inequalities in Lemmas 9, 10, and 11, we have:

$$\begin{aligned}&\mathbb{E}\Big[V(\boldsymbol{Z}(t_0+T)) - V(\boldsymbol{Z}(t_0))\Big|\boldsymbol{Z}(t_0)\Big]\\ &\leq -\theta_2\Big(\|\boldsymbol{Q}(t_0)\|_1 + \|\boldsymbol{\Psi}(t_0)\|_1\Big) + c,\ \forall t_0,\end{aligned}$$

where $c = 2c_0 + 2c_1 + c_3 + MT$.
Let $\mathcal{P} = \big\{\boldsymbol{Z} = \big(\boldsymbol{Q}, \boldsymbol{\eta}, \boldsymbol{\Psi}\big) \in \mathcal{S} : \|\boldsymbol{Q}\|_1 + \|\boldsymbol{\Psi}\|_1 \leq \frac{\bar{c}+c}{\theta_2}\big\}$ for any positive constant $\bar{c} > 0$, where $\mathcal{P}$ is a finite set of the state space $\mathcal{S}$. By this choice of $\mathcal{P}$ for the Lyapunov function $V(.)$ defined in (5.27), all the conditions of the Foster-Lyapunov theorem are satisfied, which completes the throughput optimality proof for the Blind GB-PANDAS algorithm. □

Note that the priority scheduling in the exploitation phase of the Blind GB-PANDAS algorithm is not used for the throughput optimality proof since the expected workload of a server is decreased in the same rate no matter what locality level is receiving service from the server. As long as an idle server gives service to one of the tasks in its sub-queues continuously, the system is stable. Given the routing policy, the priority scheduling is used in the exploitation phase to minimize the mean task completion time.

5.7 Simulation Results

In this section, we first compare the simulated performance of the proposed GB-PANDAS algorithm against those of Hadoop's default FCFS scheduler, Join-the-Shortest-Queue-Priority (JSQ-Priority), and JSQ-MaxWeight algorithms. Consider a computing cluster with 5000 servers where each rack consists of 50 servers and each super rack includes 10 of the racks (so four levels of locality exist). We considered geometric and log-normal distributions for processing times and under both assumptions our algorithm outperforms others. Due to the similarity of the results in the two cases we only present the results for log-normal distribution. We assumed the i-local

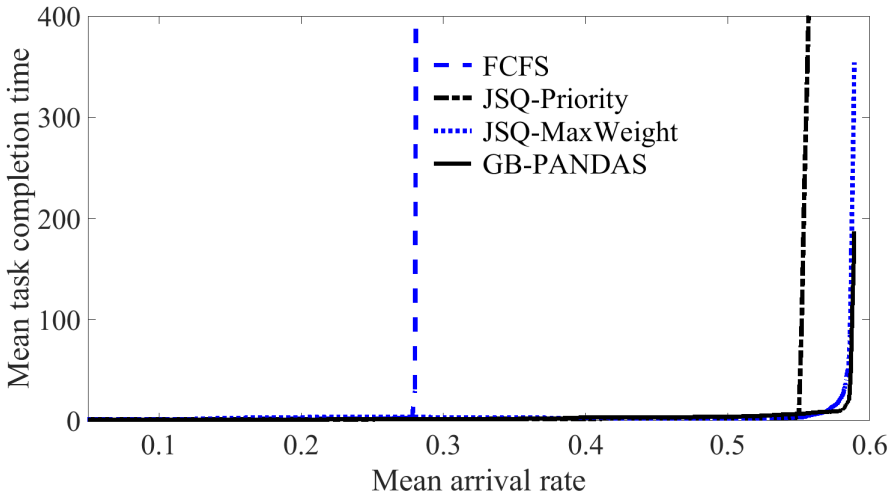

Figure 5.6: Capacity region comparison of the algorithms.

service time follows log-normal distribution with both mean and standard deviation equal to μ_i for $1 \leq i \leq 4$, where $\mu_1 = 1, \mu_2 = \frac{10}{9}, \mu_3 = \frac{5}{3}$, and $\mu_4 = 4$ (remote service is on average slower than local service by a factor of two to six times in data centers [108], and we have chosen four times slowdown in our simulations). Figure 5.6 shows the throughput performance of the four algorithms, where the y-axis shows the mean task completion time and the x-axis shows the mean arrival rate, i.e. $\frac{\sum_{\bar{l}} \lambda_{\bar{l}}}{M}$. The GB-PANDAS and JSQ-MaxWeight algorithms are throughput optimal while FCFS and JSQ-Priority algorithms are not (note that JSQ-Priority is proven to be delay optimal for two locality levels, but it is not even throughput optimal for more locality levels). Figure 5.7 compares the performance of the GB-PANDAS and JSQ-MaxWeight at high loads, where the first algorithm outperforms the latter by twofold. This significant improvement over JSQ-MaxWeight algorithm shows that JSQ-MaxWeight is not delay optimal and supports the possibility that the GB-PANDAS algorithm is delay optimal in a larger region than the JSQ-MaxWeight algorithm.

By the intuition we got from the delay optimality proof of the JSQ-MaxWeight algorithm for two locality levels in [132], [134], [201], and [204], we simulated the system under a load for which we believe JSQ-MaxWeight is delay optimal. Figure 5.8 shows the result for this specific load and we see that both the GB-PANDAS and JSQ-MaxWeight algorithms have the same performance at high loads, which again supports our guess on delay

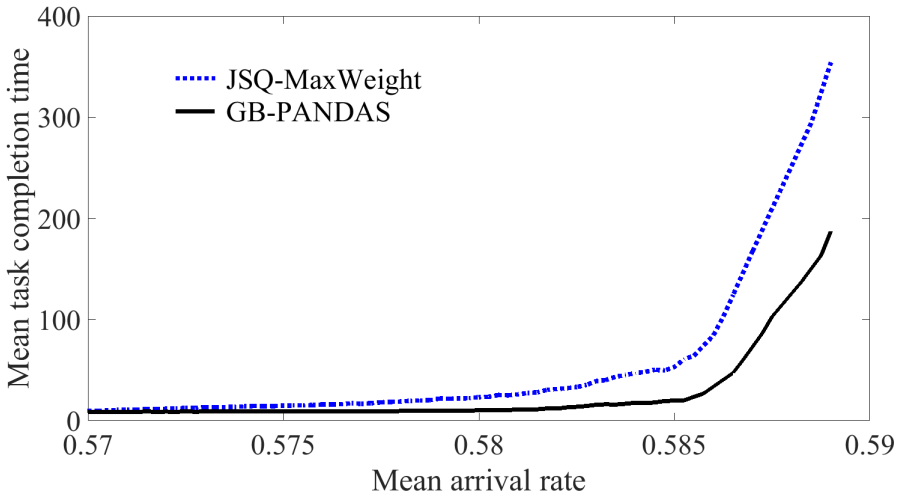

Figure 5.7: Heavy-traffic performance.

optimality of our proposed algorithm. Note that Wang et al. [132] showed that the JSQ-MaxWeight algorithm outperforms the Hadoop Fair Scheduler (HFS). Since our proposed algorithm outperforms JSQ-MaxWeight, we did not bring the HFS algorithm into our simulations.

In the following, the simulated performance of the Blind GB-PANDAS algorithm is compared with FCFS, Max-Weight, and c-μ-rule algorithms. FCFS does not use system parameters for load balancing, but Max-Weight and c-μ-rule use the same exploration-exploitation approach as Blind GB-PANDAS. Convex cost functions $C_i(Q_i) = Q_i^{1.01}$ for $i \in \{1, 2, 3\}$ are used for the c-μ-rule algorithm. Since the objective is to minimize the mean task completion time, the convexities of the cost functions are chosen so as to be close to a line for small values of Q_i. Three types of tasks and a computing cluster of three servers are considered with processing rates depicted in Figure 5.9, which are not known from the perspective of the load balancing algorithms. The task arrivals are Poisson processes with the unknown rates determined in Figure 5.9 and the processing times are log-normal that are heavy-tailed and do not have the memoryless property. Note that this affinity structure does not have the rack structure mentioned in [134] since from the processing rates of task type 2 on the three servers, servers 1 and 2 are in the same rack as server 3, but from the processing rates of task type 3 on the three servers, the second server is in the same rack as the third server, but not the first server. Hence, this affinity setup is more complicated

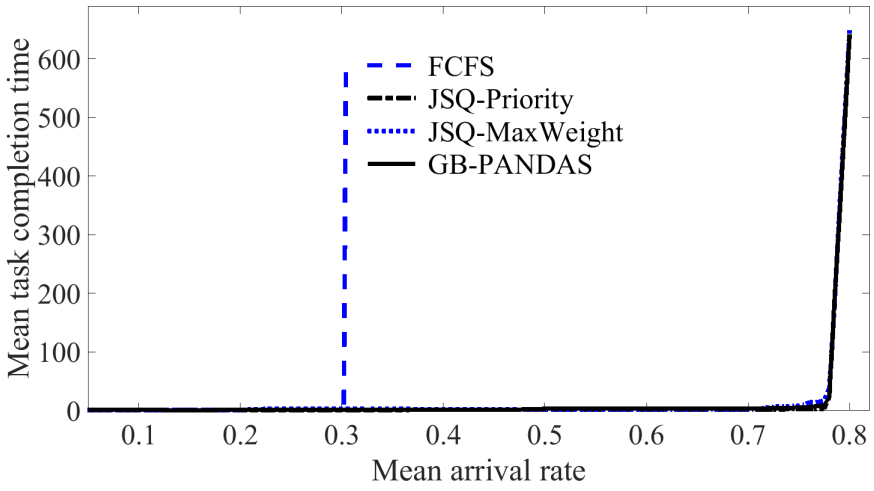

Figure 5.8: Mean task completion time under a specific load.

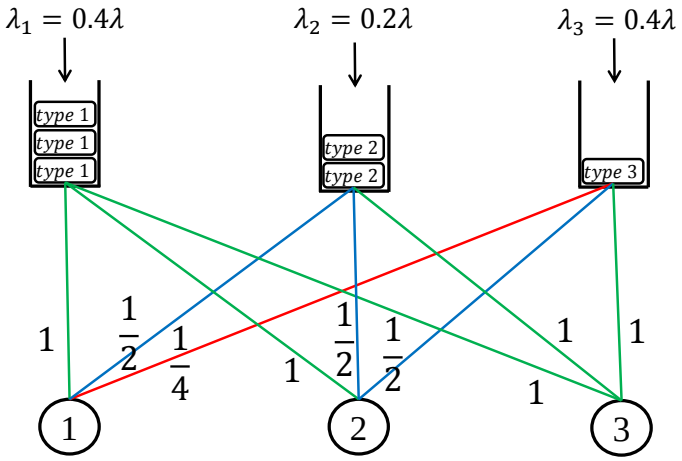

Figure 5.9: The affinity structure used for simulation with three types of tasks and three multi-skilled servers.

than the one with a rack structure.

Inspired by the fluid model planning algorithm, the following linear programming optimization should be solved to find the capacity region of the

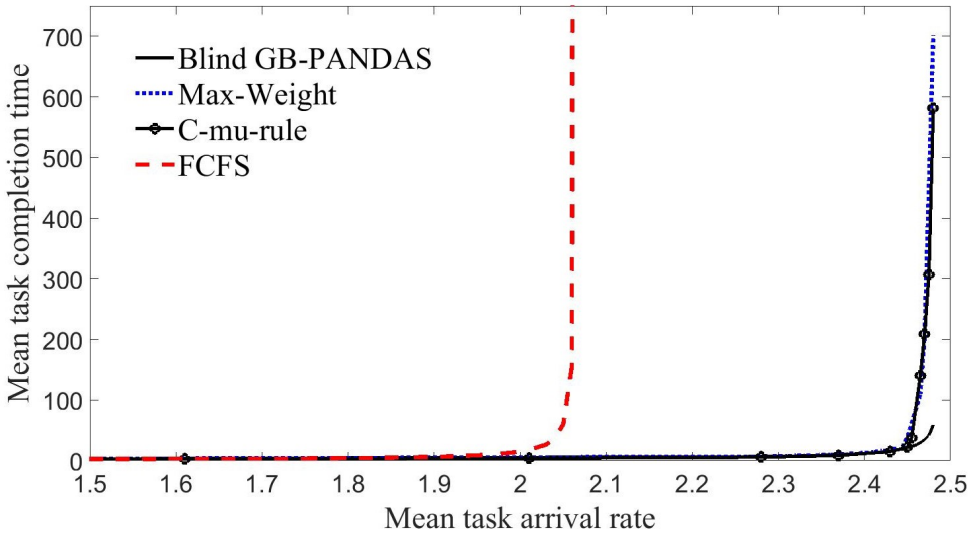

Figure 5.10: Capacity region comparison of the Blind GB-PANDAS, Max-Weight, c-μ-rule, and FCFS algorithms.

simulated system.

$$\underset{\lambda_{i,m}}{\text{maximize}} \quad \lambda = \sum_{i=1}^{3}\sum_{m=1}^{3}\lambda_{i,m}$$

subject to:

$$\begin{aligned} &\lambda_{1,1}+2\lambda_{2,1}+4\lambda_{3,1}<1, \qquad &&\lambda_{1,1}+\lambda_{1,2}+\lambda_{1,3}=0.4\lambda,\\ &\lambda_{1,2}+2\lambda_{2,2}+2\lambda_{3,2}<1, \qquad &&\lambda_{2,1}+\lambda_{2,2}+\lambda_{2,3}=0.2\lambda,\\ &\lambda_{1,3}+\lambda_{2,3}+\lambda_{3,3}<1, \qquad &&\lambda_{3,1}+\lambda_{3,2}+\lambda_{3,3}=0.4\lambda,\\ &&&\lambda_{i,m}\geq 0,\ \forall i,m\in\{1,2,3\}. \end{aligned}$$

The capacity region in terms of λ is found to be $\lambda \in [0, 2.5)$. Figure 5.10 compares the throughput performance of the four algorithms, where the mean task completion time versus the total task arrival rate, $\lambda = \sum_{i=1}^{3}\lambda_i$, is plotted. The Blind GB-PANDAS, Max-Weight, and c-μ-rule algorithms are throughput-optimal by stabilizing the system for $\lambda < 2.5$. Taking a closer look at the performance of these algorithms at high loads, Blind GB-PANDAS has a much lower mean task completion time compared to Max-Weight and c-μ-rule algorithms as depicted in Figure 5.11.

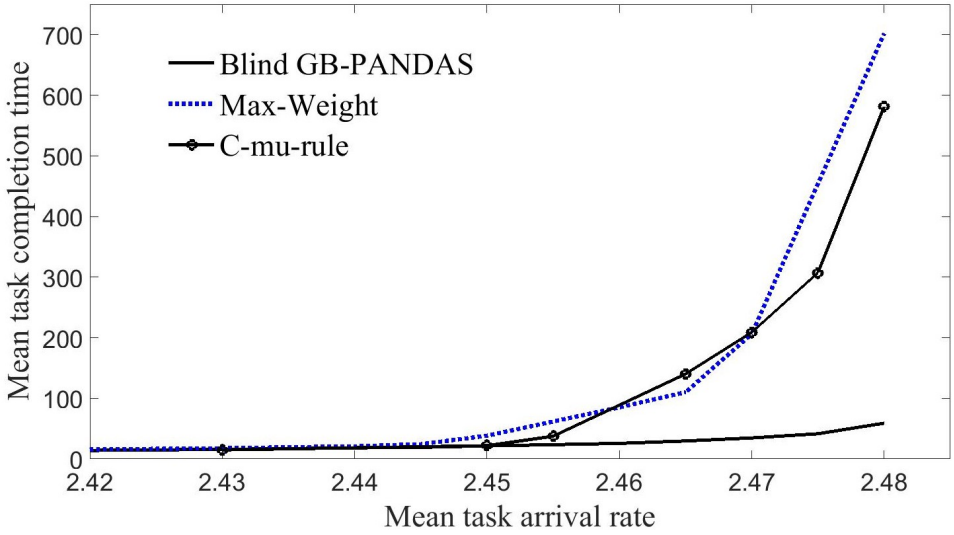

Figure 5.11: Heavy-traffic performance comparison.

Chapter 6

CONCLUSION AND DIRECTIONS FOR FUTURE RESEARCH

In this dissertation, risk-averse algorithms for multi-armed bandits, stochastic games, and stochastic congestion games are studied. The classical approaches for multi-armed bandits and games only take the expected rewards/payoffs into account. Instead, we introduce probability statements that use the reward/payoff distributions and propose a new definition of risk-averse optimality for explore-then-commit finite bandit problems and new risk-averse equilibria for stochastic games. We further used an exploration-exploitation scheme for load balancing in an affinity problem with no knowledge on task processing rates on servers and task arrival rates. The four chapters of this dissertation on explore-then-commit finite bandits, risk-averse stochastic games, stochastic congestion games, and blind load balancing are concluded in more details in the following and directions for future research are presented.

- The focus of Chapter 2 is on application domains, such as personalized health-care and one-time investment, where an experimentation phase of pure arm exploration is followed by a given finite number of exploitations of the best identified arm. We show through an example that the arm with maximum expected reward does not necessarily maximize the probability of receiving the maximum reward. We study the risk-averse explore-then-commit finite-exploitation bandits with or without considering exploration costs. In the case in which the exploration cost is not considered, we propose the OTE-MAB and FTE-MAB algorithms whose goals are to select the arm that maximizes the probability of receiving the maximum reward. We define a new notion of regret for our problem setup and find an upper bound on the minimum number of experiments that should be done to guarantee an upper bound on regret of our proposed algorithms. In the other case that the exploration cost is considered, we propose the c-OTE-MAB

algorithm for a two-armed bandit problem to determine an estimation of the optimal number of explorations. The promising future works are to introduce dynamic scores for a case where dynamic exploration is utilized in the FTE/OTE-MAB algorithms and to extend the results of the case with exploration costs to explore-then-commit multi-armed bandits with finite-time exploitations.

- We have proposed a new equilibrium for stochastic games which is risk-averse (the RAE) as a method for examining one-shot or limited-run games with stochastic payoffs in Chapter 3. In such a setting, it makes sense to consider players who want to maximize not the expected payoff but rather the likelihood of receiving the largest payoff. In doing so, we draw parallels to prospect theory in economic decisions, where consumers prefer an option with lower variance at the cost of lower expected utility, rather than an option with higher expected utility at the cost of higher variance when facing significant decisions. We then propose the risk-averse equilibrium to address one-shot games in such a situation and show it to exist in any N-player finite stochastic game. We prove the existence of the risk-averse equilibrium independent of Nash equilibrium along with familiar concepts such as strategy dominance. We also define a probability tensor and show that the risk-averse equilibria of a game are equivalent to the Nash equilibria of this tensor. We next considered the risk-averse equilibrium in limited-run games by examining M-time commit games, where players commit to a strategy for the M rounds of the stochastic game.

  Looking forward, the risk-averse equilibrium allows competition to be incorporated into many traditional risk-averse settings. Election modeling is one such example with a limited-run of interactions between candidates, and each candidate wants to maximize not the expected votes they receive, but rather their probability of winning. This is one of the drawbacks to the widely used Hotelling-Downs model [205], which assumes candidates merely want to maximize their expected voter share. By instead maximizing their probability of making the best response to other candidates, the risk-averse equilibrium will be able to address this shortcoming while offering similar interpretability and insight.

- A stochastic atomic congestion game with incomplete information on travel times along arcs of a traffic/telecommunication network is studied in Chapter 4 from a risk-averse perspective. Risk-averse travelers intend to make decisions based on probability statements regarding their travel options rather than simply taking the average travel time into account. In order to put this into perspective, we propose three classes of equilibria, i.e., risk-averse equilibrium (RAE), mean-variance equilibrium (MVE), and CVaR_α equilibrium (CVaR_αE). The MV and CVaR_α equilibria are studied in the literature for networks with simplifying assumptions such as that the probability distributions of link delays are load independent or link delays are independent, which are not the case in this work. The notions of best responses in risk-averse, mean-variance, and CVaR_α equilibria are based on maximizing the probability of traveling along the shortest path, minimizing a linear combination of mean and variance of path delay, and minimizing the expected delay at a specified risky quantile of the delay distributions, respectively. We prove that the risk-averse, mean-variance, and CVaR_α equilibria exist for any finite stochastic atomic congestion game. Although proving bounds on the price of anarchy (PoA) is not the focus of this work, we numerically study the impact of risk-averse equilibria on PoA and observe that the Braess paradox may not occur to the extent presented originally and the PoA may improve upon using any of the proposed equilibria in both Braess and Pigou networks. Promising future directions are to study non-atomic, instead of atomic, stochastic congestion games in the proposed three classes of equilibria in their general case where the arc delay distributions are load dependent and not necessarily independent of each other, to find bounds on the price of anarchy for the proposed three classes of equilibria, and to find a unified class of equilibrium that captures risk-aversion for a broader class of travel time distributions in traffic/telecommunication networks.

- The Blind GB-PANDAS algorithm is proposed in Chapter 5 for the affinity load balancing problem where no knowledge of the task arrival rates and the service rate matrix is available. An exploration-exploitation approach is proposed for load balancing which consists of exploration and exploitation phases. The system is proven to be stable

under Blind GB-PANDAS and is shown empirically through simulations to have a better delay performance than Max-Weight, c-μ-rule, and FCFS algorithms. Investigating the subspace of the capacity region in which GB-PANDAS is delay optimal is a promising direction for future work. In order to start with a simpler case, one can check whether the GB-PANDAS algorithm is heavy-traffic optimal in the same region that the JSQ-MaxWeight algorithm proposed by Wang et al. [132] is heavy-traffic optimal in a system with a nested rack structure as proposed in Chapter 5.1. Note that both GB-PANDAS and Max-Weight algorithms have high routing and scheduling computation complexity which can be alleviated using power-of-d-choices [206] or join-idle-queue [207] algorithms which are interesting directions to study as well. Another interesting future work is to consider a case where there are precedence relations between several tasks of a job, i.e. a departing task may join another queue.

Appendix A

THEOREM AND COROLLARY PROOFS OF THE OTE/FTE-MAB AND C-OTE-MAB ALGORITHMS FOR CHAPTER 2

A.1 Proof of Theorem 1

Theorem 1. For any $0 < \epsilon_r, \Delta p < 1$, if all of the K arms are experimented jointly for $N \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$ times in the experimentation phase, the one-time exploitation regret is bounded by ϵ_r, i.e. $r(\Delta p) \leq \epsilon_r$.

Proof. Consider the Bernoulli random variables $B_k = \mathbb{1}\{R_k \geq \boldsymbol{R}_{-k}\}$ and their unknown means $p_k = \mathbb{E}[B_k] = \mathbb{P}(R_k \geq \boldsymbol{R}_{-k})$ for $k \in \mathcal{K}$. Possessing N independent observations from the joint rewards of the K arms in the pure exploration phase, the confidence interval derived from Hoeffding's inequality for estimating p_k based on Equation (2.4) with confidence level $1-2e^{-\frac{a^2}{2}}$ has the property that

$$\begin{aligned}&\mathbb{P}\left(p_k \in \left(\hat{p}_k - \frac{a}{2\sqrt{N}}, \hat{p}_k + \frac{a}{2\sqrt{N}}\right)\right)\\ \geq& 1 - 2e^{-\frac{a^2}{2}},\ \forall k \in \mathcal{K}.\end{aligned} \tag{A.1}$$

In order to find a bound on regret, defined in Equation (2.5) as $r(\Delta p) = \mathbb{P}\left(p_{k^*} - p_{\hat{k}} \geq \Delta p\right)$, note that

$$\begin{aligned}&\{p_{k^*} - p_{\hat{k}} \geq \Delta p\}\\ \subseteq& \left\{\exists k \in \mathcal{K} \text{ such that } p_k \notin \left(\hat{p}_k - \frac{\Delta p}{2}, \hat{p}_k + \frac{\Delta p}{2}\right)\right\}\\ \overset{(a)}{\subseteq}& \left\{\exists k \in \mathcal{K} \text{ such that } p_k \notin \left(\hat{p}_k - \frac{a}{2\sqrt{N}}, \hat{p}_k + \frac{a}{2\sqrt{N}}\right)\right\},\end{aligned} \tag{A.2}$$

where (a) is true if $\frac{a}{2\sqrt{N}} \leq \frac{\Delta p}{2}$. By using union bound and Equation (A.1),

Portions of this appendix were previously published in Yekkehkhany et al. [176] and are used here with permission.

the probability of the right-hand side of the above equation can be bounded as follows, which results in the following bound on regret:

$$r(\Delta p) = \mathbb{P}\left(p_{k^*} - p_{\hat{k}} \geq \Delta p\right) \leq 2Ke^{-\frac{a^2}{2}} = \epsilon_r. \tag{A.3}$$

The above upper bound on regret is derived under the condition that $\frac{a}{2\sqrt{N}} \leq \frac{\Delta p}{2}$, which by using $a^2 = 2\ln\left(\frac{2K}{\epsilon_r}\right)$ and simple algebraic calculations results in $N \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$. □

A.2 Proof of Theorem 2

Theorem 2. For any $0 < \epsilon_r, \Delta p < 1$, if all of the K arms are explored jointly for N times in the experimentation phase such that $\lfloor \frac{N}{M} \rfloor \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$, the finite-time exploitation regret is bounded by ϵ_r, i.e. $r_M(\Delta p) \leq \epsilon_r$.

Proof. Consider the Bernoulli random variables $B_k^M = \mathbb{1}\{R_k^M \geq \boldsymbol{R}_{-k}^M\}$ and their unknown means $p_k^M = \mathbb{E}[B_k^M] = \mathbb{P}(R_k^M \geq \boldsymbol{R}_{-k}^M)$ for $k \in \mathcal{K}$. Possessing N independent observations from the joint rewards of the K arms in pure exploration, there are exactly $\lfloor \frac{N}{M} \rfloor$ independent samples for estimation of p_k^M. Due to the same reasoning in the proof of Theorem 1, the confidence interval for estimating p_k^M based on Equation (2.9) or (2.12) with confidence level $1 - 2e^{-\frac{a^2}{2}}$ has the property that

$$\mathbb{P}\left(p_k^M \in \left(\hat{p}_k^M - \frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}}, \hat{p}_k^M + \frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}}\right)\right) \geq 1 - 2e^{-\frac{a^2}{2}}, \tag{A.4}$$

for all $k \in \mathcal{K}$.

In order to find a bound on regret, defined in Definition 1 as $r_M(\Delta p) = \mathbb{P}\left(p_{k^*}^M - p_{\hat{k}}^M \geq \Delta p\right)$, note that

$$\begin{aligned}
&\left\{p_{k^*}^M - p_{\hat{k}}^M \geq \Delta p\right\} \\
&\subseteq \left\{\exists k \in \mathcal{K} \text{ s.t. } p_k^M \notin \left(\hat{p}_k^M - \frac{\Delta p}{2}, \hat{p}_k^M + \frac{\Delta p}{2}\right)\right\} \\
&\overset{(a)}{\subseteq} \left\{\exists k \in \mathcal{K} \text{ s.t. } p_k^M \notin \left(\hat{p}_k^M - \frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}}, \hat{p}_k^M + \frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}}\right)\right\},
\end{aligned} \tag{A.5}$$

where (a) is true if $\frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}} \leq \frac{\Delta p}{2}$. By using union bound and Equation (A.4), the probability of the right-hand side of the above equation can be bounded as follows, which results in the following bound on regret:

$$r_M(\Delta p) = \mathbb{P}\left(p_{k^*}^M - p_{\hat{k}}^M \geq \Delta p\right) \leq 2Ke^{-\frac{a^2}{2}} = \epsilon_r. \tag{A.6}$$

The above upper bound on regret is derived under the condition that $\frac{a}{2\sqrt{\lfloor \frac{N}{M} \rfloor}} \leq \frac{\Delta p}{2}$, which by using $a^2 = 2\ln\left(\frac{2K}{\epsilon_r}\right)$ and simple algebraic calculations results in $\lfloor \frac{N}{M} \rfloor \geq \frac{2\ln\left(\frac{2K}{\epsilon_r}\right)}{\Delta p^2}$. □

A.3 Proof of Theorem 3

Theorem 3. Possessing n_e number of joint experiments for the two arms and assuming that $p_{k^*} \in [0.5 + \epsilon_p, 1]$ when $\epsilon_p \in (0, 0.5]$ is an unknown parameter, we have

$$\begin{aligned} & Cr\left(\hat{N}^*(n_e), p_{k^*}\right) - Cr\left(N^*, p_{k^*}\right) \\ \leq \; & \frac{D_p}{2\sqrt{n_e}} + \Delta Cr(\hat{N}^*(n_e), n_e) \xrightarrow{n_e \to \infty} 0 \end{aligned} \tag{A.7}$$

and

$$\max_{n \in \mathcal{I}(n_e)} \left(Cr(n, p_{k^*}) - Cr(N^*, p_{k^*})\right) \leq \frac{D_p}{\sqrt{n_e}} \tag{A.8}$$

with confidence level $1 - 2e^{-\frac{a^2}{2}}$, where $\hat{N}^*(n_e)$, N^*, and $\mathcal{I}(n_e)$ are defined in Equations (2.15), (2.13), and (2.17), respectively, D_p is a constant as $D_p = \frac{a \cdot \alpha \cdot 2^{\left(4\delta_p + 1 - \frac{1}{2\ln 2}\right)}}{\sqrt{2\delta_p \ln 2}}$, where $\delta_p = \frac{1}{2}\left(-2 - \log_2(0.5 + \epsilon_p) - \log_2(0.5 - \epsilon_p)\right) > 0$, and $\Delta Cr(n, n_e) = \frac{a \cdot \alpha \cdot \sqrt{n+2} \cdot 2^{-\delta_p \cdot (n-2)}}{\sqrt{n_e}} \leq \frac{D_p}{2\sqrt{n_e}}$ for any $n \in \{1, 2, 3, \dots\}$.

Proof. The maximum deviation that $Cr_l(n, n_e)$ and $Cr_u(n, n_e)$ can have from $Cr(n, p_{k^*})$ is investigated with an associated confidence level. To this end, the maximum deviation of $r^*(n, \hat{p}_l^*(n_e))$ and $r^*(n, \hat{p}_u^*(n_e))$ from $r^*(n, p_{k^*})$ is found with the confidence level. First, the maximum deviation of $\hat{p}_l^*(n_e)$ and $\hat{p}_u^*(n_e)$ from p_{k^*} with the associated confidence level is derived below. Equation

(2.16) suggests that the following holds with confidence level $1-2e^{-\frac{a^2}{2}}$:

$$\begin{aligned} p_{k^*} - \hat{p}_l^*(n_e) &= p_{k^*} - \max\left\{\hat{p}^*(n_e) - \frac{a}{2\sqrt{n_e}}, 0.5\right\} \\ &\leq p_{k^*} - \hat{p}^*(n_e) + \frac{a}{2\sqrt{n_e}} \leq \frac{a}{2\sqrt{n_e}} + \frac{a}{2\sqrt{n_e}} = \frac{a}{\sqrt{n_e}}. \end{aligned} \tag{A.9}$$

On the other hand,

$$\begin{aligned} p_{k^*} - \hat{p}_l^*(n_e) =& p_{k^*} - \hat{p}^*(n_e) + \hat{p}^*(n_e) - \max\left\{\hat{p}^*(n_e) - \frac{a}{2\sqrt{n_e}}, 0.5\right\} \\ &\geq \max\left\{\frac{-a}{2\sqrt{n_e}}, 0.5 - \hat{p}^*(n_e)\right\} + \min\left\{\frac{a}{2\sqrt{n_e}}, \hat{p}^*(n_e) - 0.5\right\} = 0. \end{aligned} \tag{A.10}$$

The above two equations imply that $0 \leq p_{k^*} - \hat{p}_l^*(n_e) \leq \frac{a}{\sqrt{n_e}}$ with confidence level $1-2e^{-\frac{a^2}{2}}$. Similarly, it can be proved that $0 \leq \hat{p}_u^*(n_e) - p_{k^*} \leq \frac{a}{\sqrt{n_e}}$ with the mentioned confidence level.

In the following, Lipschitz constant of function $r^*(n,p)$ with respect to p is calculated by differentiating the regret function presented in Equation (2.14) with respect to p as

$$\begin{aligned} \frac{\partial r^*(n,p)}{\partial p} =& \sum_{i=\lfloor \frac{n}{2} \rfloor + 1}^{n} \binom{n}{i} \cdot (1-p)^i \cdot p^{n-i} \cdot \left(\frac{n-i}{p} - \frac{i}{1-p}\right) \\ &+ \frac{1}{2} \cdot \binom{n}{\frac{n}{2}} \cdot (1-p)^{\frac{n}{2}} \cdot p^{\frac{n}{2}} \cdot \frac{n}{2} \cdot \left(\frac{1}{p} - \frac{1}{1-p}\right) \cdot \mathbb{1}\{n \text{ is even}\}. \end{aligned} \tag{A.11}$$

Since $0.5 \leq p \leq 1$, it is easy to verify that $\frac{\partial r^*(n,p)}{\partial p} \leq 0$, so $r^*(n,p)$ is decreasing in terms of p. The derivative of $r^*(n,p)$ with respect to p calculated above can be written as follows by algebraic manipulations:

$$\frac{\partial r^*(n,p)}{\partial p} = \begin{cases} -n\binom{n-1}{\frac{n-1}{2}} p^{\frac{n-1}{2}} (1-p)^{\frac{n-1}{2}}, & \text{if } n \text{ is odd}, \\ -(n-1)\binom{n-2}{\frac{n-2}{2}} p^{\frac{n-2}{2}} (1-p)^{\frac{n-2}{2}}, & \text{if } n \text{ is even}. \end{cases} \tag{A.12}$$

Note that $\frac{\partial r^*(n,p)}{\partial p} = \frac{\partial r^*(n+1,p)}{\partial p}$ when n is an odd number and $p \in [0.5, 1]$. On the other hand, it is obvious that $r^*(n,1) = r^*(n+1,1)$, so

$$r^*(n,p) = r^*(n+1,p), \text{ if } n \text{ is odd}. \tag{A.13}$$

As a result, in terms of regret, it is not worth it to perform even number of

experiments since the last experiment does not improve regret.

It is easy to verify that $\frac{\partial r^*(n,p)}{\partial p}\big|_{p=0.5}$ can get arbitrarily large by increasing n. Hence, it is assumed that $p_{k^*} \in [0.5+\epsilon_p, 1]$, where ϵ_p can be any small number in the interval $(0, 0.5]$. In the following, the logarithm in base two of $\left|\frac{\partial r^*(n,p)}{\partial p}\right|$ is taken when n is an odd number, and as mentioned earlier, when n is even, the answer is the same as for $n-1$ which is an odd number.

$$\begin{aligned}\log_2\left|\frac{\partial r^*(n,p)}{\partial p}\right| &= \log_2 n + \log_2 \frac{(n-1)!}{\left(\left(\frac{n-1}{2}\right)!\right)^2} + \frac{n-1}{2}\Big(\log_2 p + \log_2(1-p)\Big) \\ &\overset{(a)}{\leq} \log_2 n + \Big[(n-\frac{1}{2})\log_2(n-1) - (n-1)\log_2 e + \log_2 e \\ &\quad - 2\Big(\frac{n}{2}\log_2\frac{n-1}{2} - \frac{n-1}{2}\log_2 e + \frac{1}{2}\log_2 2\pi\Big)\Big] \\ &\quad - (n-1)(1+\delta_p) \leq \frac{1}{2}\log_2(n+2) - \delta_p(n-1),\end{aligned} \tag{A.14}$$

where (a) follows by Stirling's approximation, $(n-1)! \leq (n-1)^{n-\frac{1}{2}}e^{-n+2}$ and $\left(\frac{n-1}{2}\right)! \geq \sqrt{2\pi}\left(\frac{n-1}{2}\right)^{\frac{n}{2}} e^{-\left(\frac{n-1}{2}\right)}$, and defining δ_p as follows:

$$\delta_p = \frac{1}{2}\left(-2 - \log_2(0.5+\epsilon_p) - \log_2(0.5-\epsilon_p)\right) > 0.$$

As a result,

$$\begin{aligned}&\left|\frac{\partial r^*(n,p)}{\partial p}\right| \leq \sqrt{n+2}\cdot 2^{-\delta_p(n-1)}, \\ &\lim_{n\to\infty}\left|\frac{\partial r^*(n,p)}{\partial p}\right| = 0.\end{aligned} \tag{A.15}$$

Also note that $\left|\frac{\partial r^*(n,p)}{\partial p}\right|$ given by Equation (A.12) is finite for any given n, so Equation (A.15) suggests that $\left|\frac{\partial r^*(n,p)}{\partial p}\right|$ is finite for any $n \in \{1,2,3,\dots\}$ and any $p \in [0.5+\epsilon_p, 1]$.

Equations (A.9), (A.10), (A.15), and the fact that $r^*(n,p)$ is decreasing in terms of p result in the following equation for any $n \in \{1,2,3,\dots\}$ with confidence level $1-2e^{-\frac{a^2}{2}}$:

$$\begin{aligned}0 &\leq Cr\left(n, p_{k^*}\right) - Cr_l\left(n, n_e\right) \\ &= \alpha\cdot\left[r^*\left(n, p_{k^*}\right) - r^*\left(n, \hat{p}_u^*(n_e)\right)\right] \\ &\leq \frac{a\cdot\alpha\cdot\sqrt{n+2}\cdot 2^{-\delta_p\cdot(n-1)}}{\sqrt{n_e}}.\end{aligned} \tag{A.16}$$

The above equation is true when n is odd, but recall that $r^*(n,p) = r^*(n+1,p)$ for an odd number n. In order to come up with a unified formula for $Cr\,(n,p_{k^*}) - Cr_l\,(n,n_e)$ for even and odd numbers n, define $\Delta Cr(n,n_e)$ as

$$\Delta Cr(n,n_e) \triangleq \frac{a\cdot\alpha\cdot\sqrt{n+2}\cdot 2^{-\delta_p\cdot(n-2)}}{\sqrt{n_e}}, \tag{A.17}$$

where $\lim_{n_e\to\infty} \Delta Cr(n,n_e) = 0$, $\forall n \in \{1,2,3,\cdots\}$. The same bounds can be found for $Cr_u(n,n_e) - Cr(n,p_{k^*})$, so

$$\begin{aligned} 0 \le Cr(n,p_{k^*}) - Cr_l(n,n_e) \le \Delta Cr(n,n_e),\\ 0 \le Cr_u(n,n_e) - Cr(n,p_{k^*}) \le \Delta Cr(n,n_e). \end{aligned} \tag{A.18}$$

Alternatively, the Gaussian approximation with continuity correction can be used for $r^*(n,p)$ to find an approximation for $\Delta Cr(n,n_e)$ as

$$r^*(n,p) \approx \begin{cases} \mathbb{P}\left(\tilde{r} \ge \frac{n}{2}\right), & \text{if } n \text{ is odd} \\ \frac{1}{2}\left[\mathbb{P}\left(\tilde{r} \ge \frac{n+1}{2}\right) + \mathbb{P}\left(\tilde{r} \ge \frac{n-1}{2}\right)\right], & \text{if } n \text{ is even} \end{cases}, \tag{A.19}$$

where $\tilde{r} \sim \mathcal{N}\left(n(1-p), np(1-p)\right)$. Then,

$$r^*(n,p) \approx \begin{cases} Q\left(\frac{\sqrt{n}(p-0.5)}{\sqrt{p(1-p)}}\right), & \text{if } n \text{ is odd,} \\ \frac{1}{2}\left[Q\left(\frac{\sqrt{n}(p-0.5)}{\sqrt{p(1-p)}} + \frac{1}{2\sqrt{np(1-p)}}\right) + \right. & \\ \qquad \left. Q\left(\frac{\sqrt{n}(p-0.5)}{\sqrt{p(1-p)}} - \frac{1}{2\sqrt{np(1-p)}}\right)\right], & \text{if } n \text{ is even.} \end{cases} \tag{A.20}$$

The following approximation is followed for n that is odd:

$$\frac{\partial r^*(n,p)}{\partial p} \approx \frac{-\sqrt{n}}{4\sqrt{2\pi p^3(1-p)^3}} \cdot exp\left(\frac{-n(p-0.5)^2}{2p(1-p)}\right). \tag{A.21}$$

It is easy to verify that the above approximation of $\frac{\partial r^*(n,p)}{\partial p}$ for $p \in [0.5+\epsilon_p, 1]$ approaches to zero as n goes to infinity and it is maximized at $n = \lfloor \frac{p(1-p)}{(p-0.5)^2} \rfloor$ or $n = \lceil \frac{p(1-p)}{(p-0.5)^2} \rceil$. Hence, the approximation of the partial derivative of $r^*(n,p)$ is finite for any $n \in \{1,2,3,\dots\}$. As a result, $\Delta Cr(n,n_e)$ can be

estimated by $\frac{a\cdot\alpha\cdot\sqrt{n}}{4\sqrt{2\pi p_{k^*}^3(1-p_{k^*})^3\cdot n_e}}\cdot exp\left(\frac{-n(p_{k^*}-0.5)^2}{2p_{k^*}(1-p_{k^*})}\right)$, where $p_{k^*}\in[0.5+\epsilon_p,1]$. This result is consistent with the one in Equation (A.15).

The upper bound in Equation (2.18) with confidence level $1-2e^{-\frac{a^2}{2}}$ is proved as follows. Equation (A.18) results in the following for any $n\in\{1,2,3,\dots\}$:

$$Cr\left(n,\hat{p}^*(n_e)\right)-\Delta Cr(n,n_e)\leq Cr\left(n,p_{k^*}\right)\leq Cr\left(n,\hat{p}^*(n_e)\right)+\Delta Cr(n,n_e). \tag{A.22}$$

Taking minimum from all sides of the above inequality results in

$$\begin{aligned} &Cr\left(\hat{N}^*(n_e),\hat{p}^*(n_e)\right)-\max_n\left\{\Delta Cr(n,n_e)\right\}\\ &\leq Cr\left(N^*,p_{k^*}\right)\leq Cr\left(\hat{N}^*(n_e),\hat{p}^*(n_e)\right)+\max_n\left\{\Delta Cr(n,n_e)\right\}. \end{aligned} \tag{A.23}$$

Using Equations (A.22) and (A.23) concludes as

$$\begin{aligned} &Cr\left(\hat{N}^*(n_e),p_{k^*}\right)-Cr\left(N^*,p_{k^*}\right)\\ \leq&\max_n\left\{\Delta Cr(n,n_e)\right\}+\Delta Cr(\hat{N}^*(n_e),n_e)\\ \leq&\frac{D_p}{2\sqrt{n_e}}+\Delta Cr(\hat{N}^*(n_e),n_e), \end{aligned} \tag{A.24}$$

where $D_p=\frac{a\cdot\alpha\cdot 2^{\left(4\delta_p+1-\frac{1}{2\ln 2}\right)}}{\sqrt{2\delta_p\ln 2}}$ is a constant that is derived as follows. For a given n_e, the function $\Delta Cr(n,n_e)$ is increasing in terms of n when $n<\frac{1}{2\delta_p\ln 2}-2$ and is decreasing when $n>\frac{1}{2\delta_p\ln 2}-2$. Hence, $\max_n\Delta Cr(n,n_e)\leq\Delta Cr(\frac{1}{2\delta_p\ln 2}-2,n_e)=\frac{a\cdot\alpha\cdot 2^{\left(4\delta_p-\frac{1}{2\ln 2}\right)}}{\sqrt{2\delta_p n_e\ln 2}}$.

In the following, the upper bound in Equation (2.19) with confidence level

$1 - 2e^{-\frac{a^2}{2}}$ is derived as

$$
\begin{aligned}
&\max_{n \in \mathcal{I}(n_e)} \Big(Cr(n, p_{k^*}) - Cr(N^*, p_{k^*}) \Big) \\
&\overset{(a)}{\leq} \max_{n \in \mathcal{I}(n_e)} \Big(Cr_l(n, n_e) - Cr(N^*, p_{k^*}) + \Delta Cr(n, n_e) \Big) \\
&\overset{(b)}{=} \max_{n \in \mathcal{I}(n_e)} \Big(\underbrace{Cr_l(n, n_e) - Cr_u(N_u^*, n_e)}_{\text{it is non-positive due to Equation (2.17)}} \\
&\quad + Cr_u(N_u^*, n_e) - Cr(N^*, p_{k^*}) + \Delta Cr(n, n_e) \Big) \\
&\overset{(c)}{\leq} \max_{n \in \mathcal{I}(n_e)} \Big(Cr_u(N^*, n_e) - Cr(N^*, p_{k^*}) + \Delta Cr(n, n_e) \Big) \\
&\overset{(d)}{\leq} \max_{n \in \mathcal{I}(n_e)} 2\Delta Cr(n, n_e) \leq \max_{n} 2\Delta Cr(n, n_e) \leq \frac{D_p}{\sqrt{n_e}},
\end{aligned}
\tag{A.25}
$$

where (a) follows by Equation (A.18), (b) is true by subtracting and adding the term $Cr_u(N_u^*, n_e)$, (c) uses the fact that $N_u^* = \arg\min_n Cr_u(n, n_e)$, so $Cr_u(N_u^*, n_e) \leq Cr_u(N^*, n_e)$, and (d) again follows by Equation (A.18). □

A.4 Proof of Corollary 3

Corollary 3.

$$
\lim_{n_e \to \infty} \mathbb{E}\left[\hat{N}^*(n_e)\right] = N^*. \tag{A.26}
$$

Proof. Equation (2.20) follows by the Lebesgue's Dominated Convergence Theorem since $\hat{p}^*(n_e)$ converges almost surely to p_{k^*} and $r^*(n, .)$ is positive and dominated by one half almost surely for any n_e, so $\hat{N}^*(n_e)$ is uniformly bounded by $\min\left\{n : C(m) \geq C(1) + \frac{\alpha}{2}, \forall m \geq n\right\}$ that always exists and is bounded due to the fact that $\lim_{n \to \infty} = \infty$, then

$$
\lim_{n_e \to \infty} \mathbb{E}\left[\hat{N}^*(n_e)\right] = \mathbb{E}\left[\lim_{n_e \to \infty} \hat{N}^*(n_e)\right] = \mathbb{E}\left[N^*\right] = N^*. \tag{A.27}
$$

□

A.5 Proof of Corollary 4

Corollary 4. The set of optimal stopping points N^* defined in Equation (2.13) is a subset of the set $\mathcal{I}(n_e)$ defined in Equation (2.17) with the associated confidence level, i.e. $N^* \subseteq \mathcal{I}(n_e)$ with confidence level $1 - 2e^{-\frac{a^2}{2}}$. Furthermore, $\mathcal{I}(n_e) = N^*$ with the mentioned confidence level for $n_e > \frac{D_p^2}{\left(Cr(N^*, p_{k^*}) - \min\limits_{n \notin N^*} Cr(n, p_{k^*})\right)^2}$.

Proof. The first part of the corollary is proved by contradiction. Assume by contradiction that $N^* \not\subseteq \mathcal{I}(n_e)$ with the associated confidence level, which means that

$$Cr_l\left(N^*, n_e\right) > Cr_u\left(N_u^*, n_e\right). \tag{A.28}$$

Furthermore,

$$\begin{aligned} Cr(N^*, p_{k^*}) &\geq Cr_l(N^*, n_e), \\ Cr(N_u^*, p_{k^*}) &\geq Cr(N^*, p_{k^*}), \end{aligned} \tag{A.29}$$

where the first inequality is true by Equation (A.18) and the second one is true due to the fact that N^* minimizes the function $Cr(n, p_{k^*})$. Equations (A.28) and (A.29) result in

$$Cr(N_u^*, p_{k^*}) > Cr_u\left(N_u^*, n_e\right), \tag{A.30}$$

which is a contradiction to Equation (A.18), which means that $N^* \subseteq \mathcal{I}(n_e)$ with the associated confidence level.

The second part of the corollary follows by Equation (2.19) and the fact that $N^* \subseteq \mathcal{I}(n_e)$ with the associated confidence level. If $\frac{D_p}{\sqrt{n_e}} < Cr(N^*, p_{k^*}) - \min\limits_{n \notin N^*} Cr(n, p_{k^*})$, no $n \in \{1, 2, 3, \dots\} \setminus N^*$ can satisfy Equation (2.19), but any $n \in N^*$ satisfies Equation (2.19) and $N^* \subseteq \mathcal{I}(n_e)$ which prove the second part of the corollary. □

Appendix B

THEOREM PROOF OF THE RISK-AVERSE EQUILIBRIUM FOR CHAPTER 3

B.1 Proof of Theorem 4

Theorem 4. For any finite N-player game, a risk-averse equilibrium exists.

Proof. Consider the risk-averse best response function $\boldsymbol{RB} : \boldsymbol{\Sigma} \to \boldsymbol{\Sigma}$ defined as $\boldsymbol{RB}(\boldsymbol{\sigma}) = \big(RB(\boldsymbol{\sigma}_{-1}), RB(\boldsymbol{\sigma}_{-2}), \ldots, RB(\boldsymbol{\sigma}_{-N})\big)$. The existence of a risk-averse equilibrium is equivalent to the existence of a fixed point $\boldsymbol{\sigma}^* \in \boldsymbol{\Sigma}$ such that $\boldsymbol{\sigma}^* \in \boldsymbol{RB}(\boldsymbol{\sigma}^*)$. Kakutani's Fixed Point Theorem is used to prove the existence of a fixed point for $\boldsymbol{RB}(\boldsymbol{\sigma})$. In order to use Kakutani's theorem, the four conditions listed below should be satisfied, which are proven as follows.

1. $\boldsymbol{\Sigma}$ is a nonempty subset of a finite dimensional Euclidean space, compact, and convex: $\boldsymbol{\Sigma}$ is nonempty and convex since it is the Cartesian product of nonempty simplices as each player has at least one feasible pure strategy. $\boldsymbol{\Sigma}$ is bounded since each of its elements is between zero and one, and is closed since it is the Cartesian product of simplices, so $\boldsymbol{\Sigma}$ contains all its limit points.

2. $\boldsymbol{RB}(\boldsymbol{\sigma})$ is nonempty for all $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$: $RB(\boldsymbol{\sigma}_{-i})$ is the set of all probability distributions over the set specified in Equation (3.2), where the mentioned set is nonempty since maximum always exists for finite number of values.

3. $\boldsymbol{RB}(\boldsymbol{\sigma})$ is a convex set for all $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$: It suffices to prove that $RB(\boldsymbol{\sigma}_{-i})$ is a convex set for all $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$. Consider $\sigma_i', \sigma_i'' \in RB(\boldsymbol{\sigma}_{-i})$ and $\lambda \in [0, 1]$. Define the supports of σ_i' and σ_i'' as $supp(\sigma_i') = \{s_i \in S_i : \sigma_i'(s_i) > 0\}$ and $supp(\sigma_i'') = \{s_i \in S_i : \sigma_i''(s_i) > 0\}$, respectively. From the definition of risk-averse best response in Definition 2,

$$supp(\sigma_i'), supp(\sigma_i'') \subseteq \arg\max_{s_i \in S_i} P\Big(\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\right)\Big).$$

As a result,

$$supp(\sigma_i') \cup supp(\sigma_i'') \subseteq \underset{s_i \in S_i}{\arg\max}\, P\Big(\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\right)\Big),$$

and again due to definition of risk-averse best response, any probability distribution over the set $supp(\sigma_i') \cup supp(\sigma_i'')$ is also a best response to $\boldsymbol{\sigma}_{-i}$. The mixed strategy $\lambda\sigma_i' + (1-\lambda)\sigma_i''$ is obviously a valid probability distribution over the set $supp(\sigma_i') \cup supp(\sigma_i'')$, so $\lambda\sigma_i' + (1-\lambda)\sigma_i'' \in RB(\boldsymbol{\sigma}_{-i})$ that completes the proof for convexity of the set $RB(\boldsymbol{\sigma}_{-i})$.

4. $\boldsymbol{RB(\sigma)}$ has a closed graph: $\boldsymbol{RB}(\boldsymbol{\sigma})$ has a closed graph if for any sequence $\{\boldsymbol{\sigma}^n, \widehat{\boldsymbol{\sigma}}^n\} \to \{\boldsymbol{\sigma}, \widehat{\boldsymbol{\sigma}}\}$ with $\widehat{\boldsymbol{\sigma}}^n \in \boldsymbol{RB}(\boldsymbol{\sigma}^n)$ for all $n \in \mathbb{N}$, we have $\widehat{\boldsymbol{\sigma}} \in \boldsymbol{RB}(\boldsymbol{\sigma})$. The fact that $\boldsymbol{RB}(\boldsymbol{\sigma})$ has a closed graph is proved by contradiction. Consider that $\boldsymbol{RB}(\boldsymbol{\sigma})$ does not have a closed graph. Then, there exists a sequence $\{\boldsymbol{\sigma}^n, \widehat{\boldsymbol{\sigma}}^n\} \to \{\boldsymbol{\sigma}, \widehat{\boldsymbol{\sigma}}\}$ with $\widehat{\boldsymbol{\sigma}}^n \in \boldsymbol{RB}(\boldsymbol{\sigma}^n)$ for all $n \in \mathbb{N}$, but $\widehat{\boldsymbol{\sigma}} \notin \boldsymbol{RB}(\boldsymbol{\sigma})$. This means there exists some $i \in [N]$ such that $\widehat{\sigma}_i \notin RB(\boldsymbol{\sigma}_{-i})$. As a result, due to the definition of risk-averse best response in Definition 2, there exists $\widehat{s}_i \in supp(\widehat{\sigma}_i)$, $s_i' \in S_i$, where s_i' can be any of the strategies in the set $supp(RB(\boldsymbol{\sigma}_{-i}))$, and some $\epsilon > 0$ such that

$$\begin{aligned} &P\Big(\overline{U}_i\left(s_i', \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i', \boldsymbol{\sigma}_{-i}\right)\Big) \\ >&P\Big(\overline{U}_i\left(\widehat{s}_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus \widehat{s}_i, \boldsymbol{\sigma}_{-i}\right)\Big) + 3\epsilon. \end{aligned} \tag{B.1}$$

Given that payoffs are continuous random variables and $\boldsymbol{\sigma}_{-i}^n \to \boldsymbol{\sigma}_{-i}$, for a sufficiently large n we have

$$\begin{aligned} &P\Big(\overline{U}_i\left(s_i', \boldsymbol{\sigma}_{-i}^n\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i', \boldsymbol{\sigma}_{-i}^n\right)\Big) \\ >&P\Big(\overline{U}_i\left(s_i', \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i', \boldsymbol{\sigma}_{-i}\right)\Big) - \epsilon. \end{aligned} \tag{B.2}$$

By combining Equations (B.1) and (B.2), for a sufficiently large n we have

$$\begin{aligned} &P\Big(\overline{U}_i\left(s_i', \boldsymbol{\sigma}_{-i}^n\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i', \boldsymbol{\sigma}_{-i}^n\right)\Big) \\ >&P\Big(\overline{U}_i\left(\widehat{s}_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus \widehat{s}_i, \boldsymbol{\sigma}_{-i}\right)\Big) + 2\epsilon. \end{aligned} \tag{B.3}$$

Due to the same reasoning as for Equation (B.2), for a sufficiently large

n we have

$$\begin{aligned} &P\Big(\overline{U}_i\left(\widehat{s}_i^n, \boldsymbol{\sigma}_{-i}^n\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus \widehat{s}_i^n, \boldsymbol{\sigma}_{-i}^n\right)\Big) \\ <&P\Big(\overline{U}_i\left(\widehat{s}_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus \widehat{s}_i, \boldsymbol{\sigma}_{-i}\right)\Big) + \epsilon, \end{aligned} \tag{B.4}$$

where $\widehat{s}_i^n \in supp(RB(\boldsymbol{\sigma}_{-i}^n))$. Combining Equations (B.3) and (B.4), for a sufficiently large n we have

$$\begin{aligned} &P\Big(\overline{U}_i\left(s_i', \boldsymbol{\sigma}_{-i}^n\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i', \boldsymbol{\sigma}_{-i}^n\right)\Big) \\ >&P\Big(\overline{U}_i\left(\widehat{s}_i^n, \boldsymbol{\sigma}_{-i}^n\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus \widehat{s}_i^n, \boldsymbol{\sigma}_{-i}^n\right)\Big) + \epsilon. \end{aligned} \tag{B.5}$$

However, Equation (B.5) contradicts the fact that $\widehat{s}_i^n \in supp(RB(\boldsymbol{\sigma}_{-i}^n))$.

The above four properties of the risk-averse best response function $\boldsymbol{RB}(\boldsymbol{\sigma})$ fulfil the conditions for Kakutani's Fixed Point Theorem. This means that for a finite N-player game, there always exists $\boldsymbol{\sigma}^* \in \boldsymbol{\Sigma}$ such that $\boldsymbol{\sigma}^* \in \boldsymbol{RB}(\boldsymbol{\sigma}^*)$, where by definition $\boldsymbol{\sigma}^*$ is a mixed strategy risk-averse equilibrium. □

B.2 Extra Notes on the Risk-Averse Equilibrium

The risk-averse best response of player i to the strategy profile $\boldsymbol{\sigma}_{-i}$ is presented in Definition 2 as the set of all probability distributions over the set

$$\arg\max_{s_i \in S_i} P\Big(\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\right)\Big). \tag{B.6}$$

The same randomness on the action of players $[N] \setminus i$ is considered in the random variable $\overline{U}_i(s_i, \boldsymbol{\sigma}_{-i})$ for all $s_i \in S_i$ in this work. That is why for $s_i \neq s_i' \in S_i$, the random variables $\overline{U}_i(s_i, \boldsymbol{\sigma}_{-i})$ and $\overline{U}_i(s_i', \boldsymbol{\sigma}_{-i})$ are not independent of each other in a single play of the game. On the other hand, independent randomness on the action of players $[N] \setminus i$ can be considered in $\overline{U}_i(s_i, \boldsymbol{\sigma}_{-i})$ for all $s_i \in S_i$. In that case, $\overline{U}_i(s_i, \boldsymbol{\sigma}_{-i})$ is independent from $\overline{U}_i(s_i', \boldsymbol{\sigma}_{-i})$ for

all $s_i \neq s_i' \in S_i$, so

$$
\begin{aligned}
&P\Big(\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\right)\Big) \\
&\overset{(a)}{=} \int \cdots \int_{x_{s_i} \geq \boldsymbol{x}_{S_i \setminus s_i}} \left(\prod_{s_i' \in S_i} \bar{f}_i(x_{s_i'} | (s_i', \boldsymbol{\sigma}_{-i})) dx_{s_i'} \right) \\
&\overset{(b)}{=} \int \cdots \int_{x_{s_i} \geq \boldsymbol{x}_{S_i \setminus s_i}} \left(\prod_{s_i' \in S_i} \sum_{\boldsymbol{s}_{-i} \in \boldsymbol{S}_{-i}} \left(f_i(x_{s_i'} | (s_i', \boldsymbol{s}_{-i})) \cdot \boldsymbol{\sigma}(\boldsymbol{s}_{-i}) \right) dx_{s_i'} \right) \\
&\overset{(c)}{=} \sum_{s_{-i}^{1:|S_i|} \in \boldsymbol{S}_{-i}^{|S_i|}} \left(\left(\prod_{k=1}^{|S_i|} \boldsymbol{\sigma}(\boldsymbol{s}_{-i}^k) \right) \cdot P\Big(U_i(s_i, \boldsymbol{s}_{-i}^1) \geq \boldsymbol{U}_i(S_i \setminus s_i, \boldsymbol{s}_{-i}^{2:|S_i|})\Big) \right),
\end{aligned}
\tag{B.7}
$$

where (a) follows by the fact that all payoff distributions are independent of each other, so the pdf functions can be multiplied together to get the joint distribution of $\overline{U}_i(s_i, \boldsymbol{\sigma}_{-i})$ for all $s_i \in S_i$, (b) follows by Equation (3.1), (c) is true by expanding the product and reformulating the product of the sum as the sum of products. If Equation (B.7) is used in Equation (3.6) to find the equilibrium of the game, we come up with a different equilibrium than that presented in Chapter 3. Let the equilibrium derived from Equations (3.6) and (B.7) be called RAE_2, where following the same proof of Theorem 4, RAE_2 exists for any finite N-player game. Finding a strictly dominated strategy in the framework of RAE_2 is not as straightforward as for the Nash and RAE equilibria. In the following definition, the strict dominance is described for RAE_2.

Definition 21. *A pure strategy $s_i \in S_i$ of player i strictly dominates a second pure strategy $s_i' \in S_i$ of the player if*

$$
\begin{aligned}
&P\Big(U_i\left(s_i, \boldsymbol{s}_{-i}^1\right) \geq \boldsymbol{U}_i\left(S_i \setminus s_i, \boldsymbol{s}_{-i}^{2:|S_i|}\right)\Big) \\
>&P\Big(U_i\left(s_i', \boldsymbol{s}_{-i}^1\right) \geq \boldsymbol{U}_i\left(S_i \setminus s_i', \boldsymbol{s}_{-i}^{2:|S_i|}\right)\Big), \forall \boldsymbol{s}_{-i}^{1:|S_i|} \in \boldsymbol{S}_{-i}^{|S_i|},
\end{aligned}
\tag{B.8}
$$

where what we mean by $U_i\left(s_i, \boldsymbol{s}_{-i}^1\right)$ being greater than or equal to $\boldsymbol{U}_i\Big(S_i \setminus s_i, \boldsymbol{s}_{-i}^{2:|S_i|}\Big)$ is that $U_i\left(s_i, \boldsymbol{s}_{-i}^1\right)$ is greater than or equal to $U_i\left(\widehat{s}_i, \boldsymbol{s}_{-i}^k\right)$ for all $\widehat{s}_i \in S_i \setminus s_i$, where each $\widehat{s}_i \in S_i \setminus s_i$ is associated with a possibly different pure strategy of other players $\boldsymbol{s}_{-i}^k \in \boldsymbol{S}_{-i}$ for all $2 \leq k \leq |S_i|$. Note that the associations of $\widehat{s}_i \in S_i$ and $\boldsymbol{s}_{-i}^k \in \boldsymbol{S}_{-i}$ on both sides of Equation (B.8) *remain*

the same except for s_i and s_i' for which the associations are switched with each other.

Note that the strictly dominated strategies of a player cannot be found from the risk-averse probably matrix, but finding a strictly dominated strategy needs more sophisticated calculations described in Definition 21. A strictly dominated strategy cannot be the risk-averse best response to any mixed strategy profile of other players due to the following reason. Consider that $s_i' \in S_i$ is strictly dominated by $s_i \in S_i$ for player i as is stated in Definition 21. Then, for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$, we have

$$
\begin{aligned}
& P\Big(\overline{U}_i\left(s_i, \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i, \boldsymbol{\sigma}_{-i}\right)\Big) \\
\overset{(a)}{=} & \sum_{s_{-i}^{1:|S_i|} \in \boldsymbol{S}_{-i}^{|S_i|}} \left(\left(\prod_{k=1}^{|S_i|} \boldsymbol{\sigma}(\boldsymbol{s}_{-i}^k) \right) \cdot P\Big(U_i(s_i, \boldsymbol{s}_{-i}^1) \geq \boldsymbol{U}_i(S_i \setminus s_i, \boldsymbol{s}_{-i}^{2:|S_i|})\Big) \right) \\
\overset{(b)}{>} & \sum_{s_{-i}^{1:|S_i|} \in \boldsymbol{S}_{-i}^{|S_i|}} \left(\left(\prod_{k=1}^{|S_i|} \boldsymbol{\sigma}(\boldsymbol{s}_{-i}^k) \right) \cdot P\Big(U_i(s_i', \boldsymbol{s}_{-i}^1) \geq \boldsymbol{U}_i(S_i \setminus s_i', \boldsymbol{s}_{-i}^{2:|S_i|})\Big) \right) \\
\overset{(c)}{=} & P\Big(\overline{U}_i\left(s_i', \boldsymbol{\sigma}_{-i}\right) \geq \overline{\boldsymbol{U}}_i\left(S_i \setminus s_i', \boldsymbol{\sigma}_{-i}\right)\Big),
\end{aligned}
\tag{B.9}
$$

where (a) is true by Equation (B.7), (b) is true by the assumption that the pure strategy s_i' is strictly dominated by the pure strategy s_i and using Equation (B.8) in Definition 21 on strict dominance, and (c) follows the backward direction of steps (a), (b), and (c) for pure strategy s_i' in Equation (B.7). By Equation (B.9) and Equation (3.2) in Definition 2 on the best response to a mixed strategy profile of other players, a strictly dominated pure strategy can never be a best response to any mixed strategy profile of other players. As a result, a strictly dominated pure strategy can be removed from the set of strategies of a player, and iterated elimination of strictly dominated strategies can be applied to a game under the framework of RAE_2 as well.

In order to get more insight into the new framework, the mixed strategy RAE_2 is worked out for Example 4. Consider that the first player selects U with probability σ_U and selects D otherwise. Given the first player's mixed strategy $(\sigma_U, 1-\sigma_U)$, with a little misuse of notation, denote the random variables denoting the second player's payoffs by selecting L or R with L and R, respectively. As a result, for two independent games, where in both of

them the first player independently plays according to the mixed strategy $(\sigma_U, 1-\sigma_U)$ and the second player selects L and R in the first and second games, respectively, using the law of total probability,

$$\begin{aligned} L \sim f_L(u) &= \sigma_U \cdot f_4(u) + (1-\sigma_U) \cdot f_3(u), \\ R \sim f_R(v) &= \sigma_U \cdot f_3(v) + (1-\sigma_U) \cdot f_1(v). \end{aligned} \tag{B.10}$$

The second player is indifferent between selecting L and R if $P(L \geq R) = P(R \geq L)$. Since payoffs are continuous random variables, $P(R \geq L) = 1 - P(L \geq R)$; as a result, the second player is indifferent between the strategies in two independent games if $P(L \geq R) = 0.5$. By Equation (B.10) and the fact that payoffs are independent from each other, $P(L \geq R)$ can be computed as

$$\begin{aligned} &P(L \geq R) = \int_{\infty}^{\infty} \int_{v}^{\infty} f_L(u) \cdot f_R(v)\ dudv \\ &= \int_{\infty}^{\infty} \int_{v}^{\infty} \Big(\sigma_U \cdot f_4(u) + (1-\sigma_U) \cdot f_3(u)\Big) \times \\ &\qquad \Big(\sigma_U \cdot f_3(v) + (1-\sigma_U) \cdot f_1(v)\Big)\ dudv \\ &= \sigma_U^2 \int_{\infty}^{\infty} \int_{v}^{\infty} f_4(u) \cdot f_3(v)\ dudv + \sigma_U(1-\sigma_U) \int_{\infty}^{\infty} \int_{v}^{\infty} f_4(u) \cdot f_1(v)\ dudv \\ &\quad + \sigma_U(1-\sigma_U) \int_{\infty}^{\infty} \int_{v}^{\infty} f_3(u) \cdot f_3(v)\ dudv \\ &\quad + (1-\sigma_U)^2 \int_{\infty}^{\infty} \int_{v}^{\infty} f_3(u) \cdot f_1(v)\ dudv \\ &= \sigma_U^2 P\big(U_2(U,L) \geq U_2(U,R)\big) + \sigma_U(1-\sigma_U) P\big(U_2(U,L) \geq U_2(D,R)\big) + \\ &\quad \sigma_U(1-\sigma_U) P\big(U_2(D,L) \geq U_2(U,R)\big) + (1-\sigma_U)^2 P\big(U_2(D,L) \geq U_2(D,R)\big) \\ &= \frac{2}{5}\sigma_U^2 + \sigma_U(1-\sigma_U) + \frac{1}{2}\sigma_U(1-\sigma_U) + (1-\sigma_U)^2 = -\frac{1}{10}\sigma_U^2 - \frac{1}{2}\sigma_U + 1. \end{aligned} \tag{B.11}$$

Letting $P(L \geq R) = 0.5$, then $-\frac{1}{10}\sigma_U^2 - \frac{1}{2}\sigma_U + \frac{1}{2} = 0$ whose solution is the mixed strategy RAE_2. It can be computed that $\sigma_U = \frac{-5+\sqrt{45}}{2} \approx 0.854$. As a result, due to symmetry, $\big(\sigma_1(U), \sigma_1(D)\big) = (0.854, 0.146)$ and $\big(\sigma_2(L), \sigma_2(R)\big) = (0.854, 0.146)$ form the mixed strategy RAE_2 of the game in Example 4.

The risk-averse best response under the RAE_2 framework is compared against the risk-averse best response under the RAE framework by simulation for Example 5. The mixed strategy RAE and RAE_2 exist no matter what

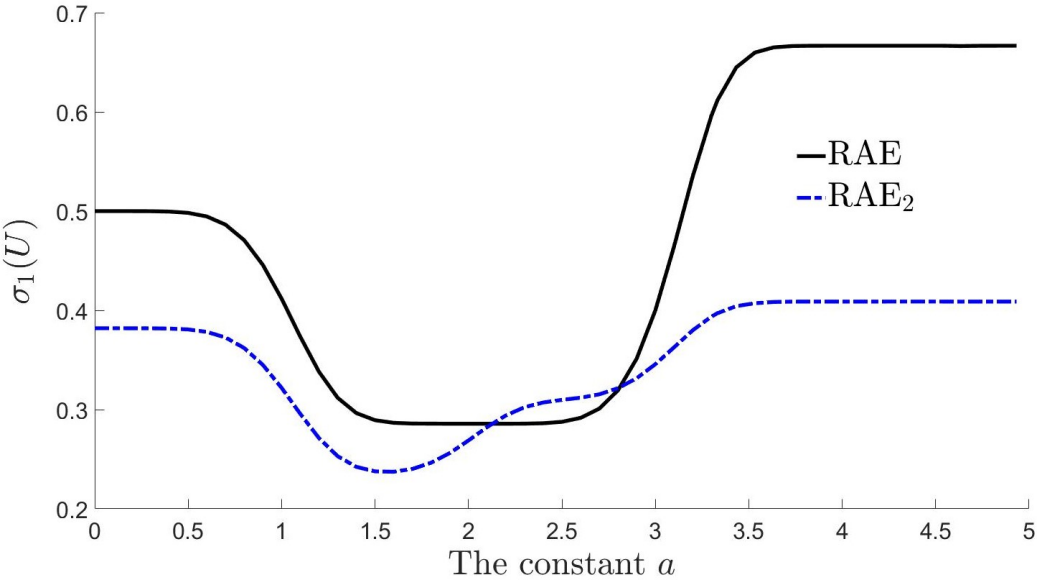

Figure B.1: The mixed strategy RAE and RAE_2 are determined by the value of $\sigma_1(U)$ in Example 5. The mixed strategies depend on the value of the constant a, where $\sigma_1(U)$ is plotted above as a function of the constant a.

the value of the positive constant a is in this example, but these equilibria depend on the value of the constant a. As described in Section 3.7, the mixed strategy RAE and RAE_2 are characterized by $\sigma_1(U)$ that is plotted in Figure B.1.

A game according to Example 5 is simulated for 10^6 rounds for a fixed constant a. In each realization of the game, the first player selects a strategy according to the mixed strategy RAE, then the payoffs of the second player for the two strategies are compared to see which is larger. After the 10^6 games, the proportion of the games in which playing strategy L outperforms playing strategy R by having a larger payoff is computed and plotted in Figure B.2 as a function of the constant a. The same procedure is performed for the mixed strategy RAE_2 and the result is plotted in the same figure. As shown in the figure, under the RAE framework, the likelihood that playing strategy L has a larger payoff than playing strategy R is the same as that of observing heads on the flip of a fair coin, which makes the second player indifferent between the two strategies. However, if the first player selects the strategy according to the mixed strategy RAE_2, it is more likely for the second player to get a larger payoff by selecting L or R except for two specific values of the constant a as shown in the figure. As a result, the risk-

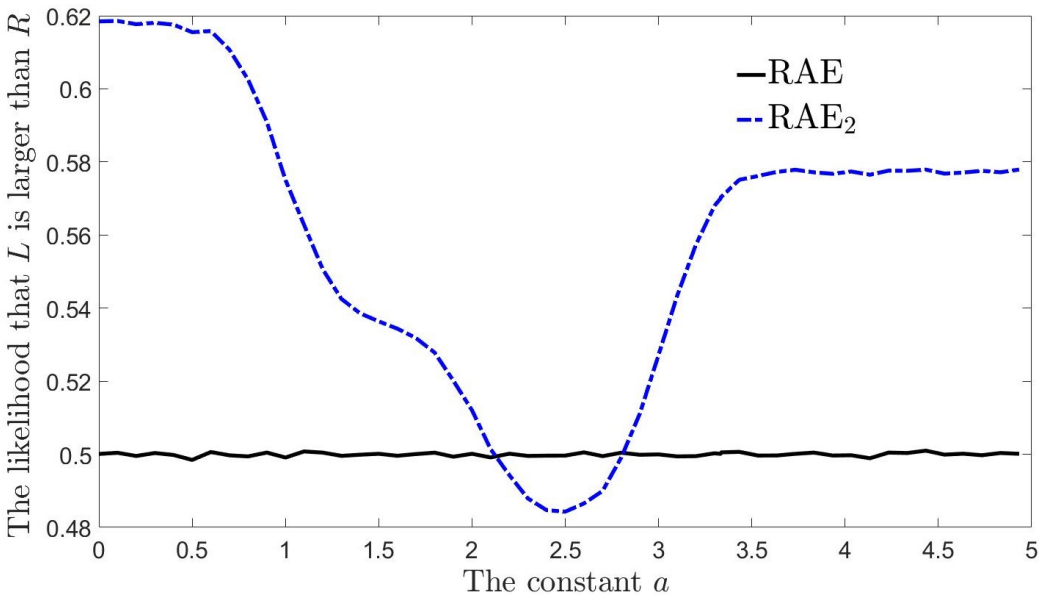

Figure B.2: The likelihood that playing strategy L outperforms playing strategy R by having a larger payoff in a single play of the game.

averse equilibrium presented in Chapter 3 makes the second player indifferent between the two strategies since the chances of receiving a larger payoff from either strategy are the same in a single play of the game.

Appendix C

THEOREM PROOFS OF THE RISK-AVERSE EQUILIBRIUM FOR CHAPTER 4

C.1 Proof of Theorem 5

Theorem 5. For any finite n-player stochastic congestion game, a risk-averse equilibrium exists.

Proof. Let $\boldsymbol{RB} : \boldsymbol{\Sigma} \to \boldsymbol{\Sigma}$ be the risk-averse best response function where $\boldsymbol{RB}(\boldsymbol{\sigma}) = \big(RB(\boldsymbol{\sigma}_{-1}), RB(\boldsymbol{\sigma}_{-2}), \ldots, RB(\boldsymbol{\sigma}_{-N})\big)$. It is easy to see that the existence of a fixed point $\boldsymbol{\sigma}^* \in \boldsymbol{\Sigma}$ for the risk-averse best response function, i.e., $\boldsymbol{\sigma}^* \in \boldsymbol{RB}(\boldsymbol{\sigma}^*)$, proves the existence of a risk-averse equilibrium. The following four conditions of the Kakutani's Fixed Point Theorem are shown to be satisfied for the function $\boldsymbol{RB}(\boldsymbol{\sigma})$ to prove the existence of a fixed point for the function.

1. The domain of function $\boldsymbol{RB}(.)$ is a non-empty, compact, and convex subset of a finite dimensional Euclidean space: $\boldsymbol{\Sigma}$ is the Cartesian product of non-empty simplices as each player has at least one strategy to play; furthermore, each of the elements of $\boldsymbol{\Sigma}$ is between zero and one, so $\boldsymbol{\Sigma}$ is non-empty, convex, bounded, and closed containing all its limit points.

2. $\boldsymbol{RB(\sigma)} \neq \varnothing$, $\forall \boldsymbol{\sigma} \in \boldsymbol{\Sigma}$: The set in Equation (4.3) is non-empty as maximum exists over a finite number of values. As a result, $RB(\boldsymbol{\sigma}_{-i})$ is non-empty for all $i \in [n]$ since it is the set of all probability distributions over the corresponding mentioned non-empty set.

3. The co-domain of function $\boldsymbol{RB}(.)$ is a convex set for all $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$: It suffices to prove that $RB(\boldsymbol{\sigma}_{-i})$ is a convex set for all $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$ and for all $i \in [n]$. For any $i \in [n]$, if $\sigma_i, \sigma_i' \in RB(\boldsymbol{\sigma}_{-i})$, we need to prove that $\lambda\sigma_i + (1-\lambda)\sigma_i' \in RB(\boldsymbol{\sigma}_{-i})$ for any $\lambda \in [0,1]$ and for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$. Let the supports of σ_i and σ_i' be defined as $supp(\sigma_i) = \{p_i \in \mathcal{P}_i :$

$\sigma_i(p_i) > 0\}$ and $supp(\sigma'_i) = \{p_i \in \mathcal{P}_i : \sigma'_i(p_i) > 0\}$, respectively. It is concluded from the definition of the risk-averse best response in Definition 11 that $supp(\sigma_i), supp(\sigma'_i) \subseteq \underset{p_i \in \mathcal{P}_i}{\arg\max}\ P\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p_i, \boldsymbol{\sigma}_{-i})\right)$, which results in

$$supp(\sigma_i) \cup supp(\sigma'_i) \subseteq \underset{p_i \in \mathcal{P}_i}{\arg\max}\ P\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p_i, \boldsymbol{\sigma}_{-i})\right).$$

As a result, using the definition of risk-averse best response, any probability distribution over the set $supp(\sigma_i) \cup supp(\sigma'_i)$ is a risk-averse best response to $\boldsymbol{\sigma}_{-i}$. It is trivial that the mixed strategy $\lambda\sigma_i + (1-\lambda)\sigma'_i$ is a valid probability distribution over the set $supp(\sigma_i) \cup supp(\sigma'_i)$ for any $\lambda \in [0,1]$, so $\lambda\sigma_i + (1-\lambda)\sigma'_i \in RB(\boldsymbol{\sigma}_{-i})$ for any $\lambda \in [0,1]$ and for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$ that completes the convexity proof of the set $RB(\boldsymbol{\sigma}_{-i})$.

4. $\boldsymbol{RB}(\boldsymbol{\sigma})$ has a closed graph: $\boldsymbol{RB}(\boldsymbol{\sigma})$ has a closed graph if for any sequence $\{\boldsymbol{\sigma}^m, \widehat{\boldsymbol{\sigma}}^m\} \to \{\boldsymbol{\sigma}, \widehat{\boldsymbol{\sigma}}\}$ with $\widehat{\boldsymbol{\sigma}}^m \in \boldsymbol{RB}(\boldsymbol{\sigma}^m)$ for all $m \in \mathbb{N}$, we have $\widehat{\boldsymbol{\sigma}} \in \boldsymbol{RB}(\boldsymbol{\sigma})$. Proof by contradiction is used to show that $\boldsymbol{RB}(\boldsymbol{\sigma})$ has a closed graph. Consider by contradiction that $\boldsymbol{RB}(\boldsymbol{\sigma})$ does not have a closed graph, so there exists a sequence $\{\boldsymbol{\sigma}^m, \widehat{\boldsymbol{\sigma}}^m\} \to \{\boldsymbol{\sigma}, \widehat{\boldsymbol{\sigma}}\}$ with $\widehat{\boldsymbol{\sigma}}^m \in \boldsymbol{RB}(\boldsymbol{\sigma}^m)$ for all $m \in \mathbb{N}$, but $\widehat{\boldsymbol{\sigma}} \notin \boldsymbol{RB}(\boldsymbol{\sigma})$. As a result, there exists some $i \in [n]$ such that $\widehat{\sigma}_i \notin RB(\boldsymbol{\sigma}_{-i})$. Using the definition of risk-averse best response in Definition 11, there exists $p'_i \in supp(RB(\boldsymbol{\sigma}_{-i}))$, $\widehat{p}_i \in supp(\widehat{\sigma}_i)$, and some $\epsilon > 0$ such that

$$\begin{aligned} &P\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p'_i, \boldsymbol{\sigma}_{-i})\right) \\ >&P\left(\overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus \widehat{p}_i, \boldsymbol{\sigma}_{-i})\right) + 3\epsilon. \end{aligned} \tag{C.1}$$

Since the latencies over edges are continuous random variables and $\boldsymbol{\sigma}^m_{-i} \to \boldsymbol{\sigma}_{-i}$, for any $\epsilon > 0$, there exists a sufficiently large m_1 such that we have the following for $m \geq m_1$:

$$\begin{aligned} &P\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}^m_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p'_i, \boldsymbol{\sigma}^m_{-i})\right) \\ >&P\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p'_i, \boldsymbol{\sigma}_{-i})\right) - \epsilon. \end{aligned} \tag{C.2}$$

By adding inequalities with the same direction in Equations (C.1) and

(C.2), for $m \geq m_1$ we have

$$\begin{aligned} & P\left(\overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}^m) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p_i', \boldsymbol{\sigma}_{-i}^m)\right) \\ >& P\left(\overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus \widehat{p}_i, \boldsymbol{\sigma}_{-i})\right) + 2\epsilon. \end{aligned} \tag{C.3}$$

For the same reason as of Equation (C.2), for any $\epsilon > 0$, there exists a sufficiently large m_2 such that we have the following for $m \geq m_2$:

$$\begin{aligned} & P\left(\overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus \widehat{p}_i, \boldsymbol{\sigma}_{-i})\right) \\ >& P\left(\overline{L}^i(\widehat{p}_i^m, \boldsymbol{\sigma}_{-i}^m) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus \widehat{p}_i^m, \boldsymbol{\sigma}_{-i}^m)\right) - \epsilon, \end{aligned} \tag{C.4}$$

where $\widehat{p}_i^m \in supp(RB(\boldsymbol{\sigma}_{-i}^m))$. By adding the inequalities with the same direction in Equations (C.3) and (C.4), for $m \geq \max\{m_1, m_2\}$ we have

$$\begin{aligned} & P\left(\overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}^m) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus p_i', \boldsymbol{\sigma}_{-i}^m)\right) \\ >& P\left(\overline{L}^i(\widehat{p}_i^m, \boldsymbol{\sigma}_{-i}^m) \leq \overline{\boldsymbol{L}}^i(\mathcal{P}_i \setminus \widehat{p}_i^m, \boldsymbol{\sigma}_{-i}^m)\right) + \epsilon. \end{aligned} \tag{C.5}$$

Equation (C.5) contradicts the fact that $\widehat{p}_i^m \in supp(RB(\boldsymbol{\sigma}_{-i}^m))$, which completes the proof that $\boldsymbol{RB}(\boldsymbol{\sigma})$ has a closed graph.

As listed above, the risk-averse best response function $\boldsymbol{RB}(\boldsymbol{\sigma})$ satisfies the four conditions of Kakutani's Fixed Point Theorem. As a direct result, for any finite n-player stochastic congestion game, there exists $\boldsymbol{\sigma}^* \in \boldsymbol{\Sigma}$ such that $\boldsymbol{\sigma}^* \in \boldsymbol{RB}(\boldsymbol{\sigma}^*)$, which completes the existence proof of a risk-averse equilibrium for such games. □

C.2 Proof of Theorem 6

Theorem 6. For any finite n-player stochastic congestion game, a mean-variance equilibrium exists.

Proof. Let $\boldsymbol{MB} : \boldsymbol{\Sigma} \to \boldsymbol{\Sigma}$ be the mean-variance best response function where $\boldsymbol{MB}(\boldsymbol{\sigma}) = \big(MB(\boldsymbol{\sigma}_{-1}), MB(\boldsymbol{\sigma}_{-2}), \ldots, MB(\boldsymbol{\sigma}_{-N})\big)$. It is easy to see that the existence of a fixed point $\boldsymbol{\sigma}^* \in \boldsymbol{\Sigma}$ for the mean-variance best response function, i.e., $\boldsymbol{\sigma}^* \in \boldsymbol{MB}(\boldsymbol{\sigma}^*)$, proves the existence of a mean-variance equilibrium. The following four conditions of the Kakutani's Fixed Point

Theorem are shown to be satisfied for the function $\boldsymbol{MB}(\boldsymbol{\sigma})$ to prove the existence of a fixed point for the function.

1. The domain of function $\boldsymbol{MB}(.)$ is a non-empty, compact, and convex subset of a finite dimensional Euclidean space: $\boldsymbol{\Sigma}$ is the Cartesian product of non-empty simplices as each player has at least one strategy to play; furthermore, each of the elements of $\boldsymbol{\Sigma}$ is between zero and one, so $\boldsymbol{\Sigma}$ is non-empty, convex, bounded, and closed containing all its limit points.

2. $\boldsymbol{MB}(\boldsymbol{\sigma}) \neq \varnothing$, $\forall \boldsymbol{\sigma} \in \boldsymbol{\Sigma}$: The set in Equation (4.9) is non-empty as minimum exists over a finite number of values. As a result, $MB(\boldsymbol{\sigma}_{-i})$ is non-empty for all $i \in [n]$ since it is the set of all probability distributions over the corresponding mentioned non-empty set.

3. The co-domain of function $\boldsymbol{MB}(.)$ is a convex set for all $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$: It suffices to prove that $MB(\boldsymbol{\sigma}_{-i})$ is a convex set for all $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$ and for all $i \in [n]$. For any $i \in [n]$, if $\sigma_i, \sigma_i' \in MB(\boldsymbol{\sigma}_{-i})$, we need to prove that $\lambda\sigma_i + (1-\lambda)\sigma_i' \in MB(\boldsymbol{\sigma}_{-i})$ for any $\lambda \in [0,1]$ and for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$. Let the supports of σ_i and σ_i' be defined as $supp(\sigma_i) = \{p_i \in \mathcal{P}_i : \sigma_i(p_i) > 0\}$ and $supp(\sigma_i') = \{p_i \in \mathcal{P}_i : \sigma_i'(p_i) > 0\}$, respectively. It is concluded from the definition of the mean-variance best response in Definition 14 that $supp(\sigma_i), supp(\sigma_i') \subseteq \underset{p_i \in \mathcal{P}_i}{\arg\min}\ \text{Var}\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p_i, \boldsymbol{\sigma}_{-i})$, which results in

$$supp(\sigma_i) \cup supp(\sigma_i') \subseteq \underset{p_i \in \mathcal{P}_i}{\arg\min}\ \text{Var}\left(\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p_i, \boldsymbol{\sigma}_{-i}).$$

   As a result, using the definition of mean-variance best response, any probability distribution over the set $supp(\sigma_i) \cup supp(\sigma_i')$ is a mean-variance best response to $\boldsymbol{\sigma}_{-i}$. The mixed strategy $\lambda\sigma_i + (1-\lambda)\sigma_i'$ is obviously a valid probability distribution over the set $supp(\sigma_i) \cup supp(\sigma_i')$ for any $\lambda \in [0,1]$, so $\lambda\sigma_i + (1-\lambda)\sigma_i' \in MB(\boldsymbol{\sigma}_{-i})$ for any $\lambda \in [0,1]$ and for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$ that completes the convexity proof of the set $MB(\boldsymbol{\sigma}_{-i})$.

4. $\boldsymbol{MB}(\boldsymbol{\sigma})$ has a closed graph: $\boldsymbol{MB}(\boldsymbol{\sigma})$ has a closed graph if for any sequence $\{\boldsymbol{\sigma}^m, \widehat{\boldsymbol{\sigma}}^m\} \to \{\boldsymbol{\sigma}, \widehat{\boldsymbol{\sigma}}\}$ with $\widehat{\boldsymbol{\sigma}}^m \in \boldsymbol{MB}(\boldsymbol{\sigma}^m)$ for all $m \in \mathbb{N}$, we have $\widehat{\boldsymbol{\sigma}} \in \boldsymbol{MB}(\boldsymbol{\sigma})$. Proof by contradiction is used to show that $\boldsymbol{MB}(\boldsymbol{\sigma})$ has a closed graph. Consider by contradiction that $\boldsymbol{MB}(\boldsymbol{\sigma})$ does not have a closed graph, so there exists a sequence $\{\boldsymbol{\sigma}^m, \widehat{\boldsymbol{\sigma}}^m\} \to \{\boldsymbol{\sigma}, \widehat{\boldsymbol{\sigma}}\}$ with

$\widehat{\boldsymbol{\sigma}}^m \in \boldsymbol{MB}(\boldsymbol{\sigma}^m)$ for all $m \in \mathbb{N}$, but $\widehat{\boldsymbol{\sigma}} \notin \boldsymbol{MB}(\boldsymbol{\sigma})$. As a result, there exists some $i \in [n]$ such that $\widehat{\sigma}_i \notin MB(\boldsymbol{\sigma}_{-i})$. Using the definition of mean-variance best response in Definition 14, there exists $p'_i \in supp(MB(\boldsymbol{\sigma}_{-i}))$, $\widehat{p}_i \in supp(\widehat{\sigma}_i)$, and some $\epsilon > 0$ such that

$$\begin{aligned} &\text{Var}\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p'_i, \boldsymbol{\sigma}_{-i}) \\ <&\text{Var}\left(\overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) - 3\epsilon. \end{aligned} \tag{C.6}$$

Since the latencies over edges are continuous random variables and $\boldsymbol{\sigma}^m_{-i} \to \boldsymbol{\sigma}_{-i}$, for any $\epsilon > 0$, there exists a sufficiently large m_3 such that we have the following for $m \geq m_3$:

$$\begin{aligned} &\text{Var}\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}^m_{-i})\right) + \rho \cdot \overline{l}^i(p'_i, \boldsymbol{\sigma}^m_{-i}) \\ <&\text{Var}\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(p'_i, \boldsymbol{\sigma}_{-i}) + \epsilon. \end{aligned} \tag{C.7}$$

By adding inequalities with the same direction in Equations (C.6) and (C.7), for $m \geq m_3$ we have

$$\begin{aligned} &\text{Var}\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}^m_{-i})\right) + \rho \cdot \overline{l}^i(p'_i, \boldsymbol{\sigma}^m_{-i}) \\ <&\text{Var}\left(\overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) - 2\epsilon. \end{aligned} \tag{C.8}$$

For the same reason as of Equation (C.7), for any $\epsilon > 0$, there exists a sufficiently large m_4 such that we have the following for $m \geq m_4$:

$$\begin{aligned} &\text{Var}\left(\overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i})\right) + \rho \cdot \overline{l}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) \\ <&\text{Var}\left(\overline{L}^i(\widehat{p}^m_i, \boldsymbol{\sigma}^m_{-i})\right) + \rho \cdot \overline{l}^i(\widehat{p}^m_i, \boldsymbol{\sigma}^m_{-i}) + \epsilon, \end{aligned} \tag{C.9}$$

where $\widehat{p}^m_i \in supp(MB(\boldsymbol{\sigma}^m_{-i}))$. By adding the inequalities with the same direction in Equations (C.8) and (C.9), for $m \geq \max\{m_3, m_4\}$ we have

$$\begin{aligned} &\text{Var}\left(\overline{L}^i(p'_i, \boldsymbol{\sigma}^m_{-i})\right) + \rho \cdot \overline{l}^i(p'_i, \boldsymbol{\sigma}^m_{-i}) \\ <&\text{Var}\left(\overline{L}^i(\widehat{p}^m_i, \boldsymbol{\sigma}^m_{-i})\right) + \rho \cdot \overline{l}^i(\widehat{p}^m_i, \boldsymbol{\sigma}^m_{-i}) - \epsilon. \end{aligned} \tag{C.10}$$

Equation (C.10) contradicts the fact that $\widehat{p}^m_i \in supp(MB(\boldsymbol{\sigma}^m_{-i}))$, which completes the proof that $\boldsymbol{MB}(\boldsymbol{\sigma})$ has a closed graph.

As listed above, the mean-variance best response function $\boldsymbol{MB}(\boldsymbol{\sigma})$ satisfies the four conditions of Kakutani's Fixed Point Theorem. As a direct result, for any finite n-player stochastic congestion game, there exists $\boldsymbol{\sigma}^* \in \boldsymbol{\Sigma}$ such that $\boldsymbol{\sigma}^* \in \boldsymbol{MB}(\boldsymbol{\sigma}^*)$, which completes the existence proof of a mean-variance equilibrium for such games. □

C.3 Proof of Theorem 7

Theorem 7. For any finite n-player stochastic congestion game, a CVaR_α equilibrium exists.

Proof. Let $\boldsymbol{CB} : \boldsymbol{\Sigma} \to \boldsymbol{\Sigma}$ be the CVaR_α best response function where $\boldsymbol{CB}(\boldsymbol{\sigma}) = \big(CB(\boldsymbol{\sigma}_{-1}), CB(\boldsymbol{\sigma}_{-2}), \ldots, CB(\boldsymbol{\sigma}_{-N})\big)$. It is easy to see that the existence of a fixed point $\boldsymbol{\sigma}^* \in \boldsymbol{\Sigma}$ for the CVaR_α best response function, i.e., $\boldsymbol{\sigma}^* \in \boldsymbol{CB}(\boldsymbol{\sigma}^*)$, proves the existence of a CVaR_α equilibrium. The following four conditions of the Kakutani's Fixed Point Theorem are shown to be satisfied for the function $\boldsymbol{CB}(\boldsymbol{\sigma})$ to prove the existence of a fixed point for the function.

1. The domain of function $\boldsymbol{CB}(.)$ is a non-empty, compact, and convex subset of a finite dimensional Euclidean space: $\boldsymbol{\Sigma}$ is the Cartesian product of non-empty simplices as each player has at least one strategy to play; furthermore, each of the elements of $\boldsymbol{\Sigma}$ is between zero and one, so $\boldsymbol{\Sigma}$ is non-empty, convex, bounded, and closed containing all its limit points.

2. $\boldsymbol{CB(\sigma)} \neq \varnothing$, $\forall \boldsymbol{\sigma} \in \boldsymbol{\Sigma}$: The set in Equation (4.19) is non-empty as minimum exists over a finite number of values. As a result, $CB(\boldsymbol{\sigma}_{-i})$ is non-empty for all $i \in [n]$ since it is the set of all probability distributions over the corresponding mentioned non-empty set.

3. The co-domain of function $\boldsymbol{CB}(.)$ is a convex set for all $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$: It suffices to prove that $CB(\boldsymbol{\sigma}_{-i})$ is a convex set for all $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$ and for all $i \in [n]$. For any $i \in [n]$, if $\sigma_i, \sigma_i' \in CB(\boldsymbol{\sigma}_{-i})$, we need to prove that $\lambda\sigma_i + (1-\lambda)\sigma_i' \in CB(\boldsymbol{\sigma}_{-i})$ for any $\lambda \in [0,1]$ and for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$. Let the supports of σ_i and σ_i' be defined as $supp(\sigma_i) = \{p_i \in \mathcal{P}_i : \sigma_i(p_i) > 0\}$ and $supp(\sigma_i') = \{p_i \in \mathcal{P}_i : \sigma_i'(p_i) > 0\}$, respectively. It is concluded from the definition

of the CVaR_α best response in Definition 17 that $supp(\sigma_i), supp(\sigma_i') \subseteq \arg\min_{p_i \in \mathcal{P}_i} \text{E}\left[\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\right]$, which results in

$$supp(\sigma_i) \cup supp(\sigma_i') \subseteq \arg\min_{p_i \in \mathcal{P}_i} \text{E}\left[\overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p_i, \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(p_i, \boldsymbol{\sigma}_{-i})\right].$$

As a result, using the definition of CVaR_α best response, any probability distribution over the set $supp(\sigma_i) \cup supp(\sigma_i')$ is a CVaR_α best response to $\boldsymbol{\sigma}_{-i}$. The mixed strategy $\lambda\sigma_i + (1-\lambda)\sigma_i'$ is obviously a valid probability distribution over the set $supp(\sigma_i) \cup supp(\sigma_i')$ for any $\lambda \in [0,1]$, so $\lambda\sigma_i + (1-\lambda)\sigma_i' \in CB(\boldsymbol{\sigma}_{-i})$ for any $\lambda \in [0,1]$ and for any $\boldsymbol{\sigma}_{-i} \in \boldsymbol{\Sigma}_{-i}$ that completes the convexity proof of the set $CB(\boldsymbol{\sigma}_{-i})$.

4. $\boldsymbol{CB}(\boldsymbol{\sigma})$ has a closed graph: $\boldsymbol{CB}(\boldsymbol{\sigma})$ has a closed graph if for any sequence $\{\boldsymbol{\sigma}^m, \widehat{\boldsymbol{\sigma}}^m\} \to \{\boldsymbol{\sigma}, \widehat{\boldsymbol{\sigma}}\}$ with $\widehat{\boldsymbol{\sigma}}^m \in \boldsymbol{CB}(\boldsymbol{\sigma}^m)$ for all $m \in \mathbb{N}$, we have $\widehat{\boldsymbol{\sigma}} \in \boldsymbol{CB}(\boldsymbol{\sigma})$. Proof by contradiction is used to show that $\boldsymbol{CB}(\boldsymbol{\sigma})$ has a closed graph. Consider by contradiction that $\boldsymbol{CB}(\boldsymbol{\sigma})$ does not have a closed graph, so there exists a sequence $\{\boldsymbol{\sigma}^m, \widehat{\boldsymbol{\sigma}}^m\} \to \{\boldsymbol{\sigma}, \widehat{\boldsymbol{\sigma}}\}$ with $\widehat{\boldsymbol{\sigma}}^m \in \boldsymbol{CB}(\boldsymbol{\sigma}^m)$ for all $m \in \mathbb{N}$, but $\widehat{\boldsymbol{\sigma}} \notin \boldsymbol{CB}(\boldsymbol{\sigma})$. As a result, there exists some $i \in [n]$ such that $\widehat{\sigma}_i \notin CB(\boldsymbol{\sigma}_{-i})$. Using the definition of CVaR_α best response in Definition 17, there exists $p_i' \in supp(CB(\boldsymbol{\sigma}_{-i}))$, $\widehat{p}_i \in supp(\widehat{\sigma}_i)$, and some $\epsilon > 0$ such that

$$\begin{aligned} &\text{E}\left[\overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(p_i', \boldsymbol{\sigma}_{-i})\right] \\ <&\text{E}\left[\overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i})\right] - 3\epsilon. \end{aligned} \tag{C.11}$$

Since the latencies over edges are continuous random variables and $\boldsymbol{\sigma}_{-i}^m \to \boldsymbol{\sigma}_{-i}$, for any $\epsilon > 0$, there exists a sufficiently large m_5 such that we have the following for $m \geq m_5$:

$$\begin{aligned} &\text{E}\left[\overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}^m) \middle| \overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}^m) \geq v_\alpha^i(p_i', \boldsymbol{\sigma}_{-i}^m)\right] \\ <&\text{E}\left[\overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(p_i', \boldsymbol{\sigma}_{-i})\right] + \epsilon. \end{aligned} \tag{C.12}$$

By adding inequalities with the same direction in Equations (C.11) and

(C.12), for $m \geq m_5$ we have

$$\begin{aligned} &\mathrm{E}\left[\overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}^m) \middle| \overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}^m) \geq v_\alpha^i(p_i', \boldsymbol{\sigma}_{-i}^m)\right] \\ <&\mathrm{E}\left[\overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i})\right] - 2\epsilon. \end{aligned} \tag{C.13}$$

For the same reason as of Equation (C.12), for any $\epsilon > 0$, there exists a sufficiently large m_6 such that we have the following for $m \geq m_6$:

$$\begin{aligned} &\mathrm{E}\left[\overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) \middle| \overline{L}^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i}) \geq v_\alpha^i(\widehat{p}_i, \boldsymbol{\sigma}_{-i})\right] \\ <&\mathrm{E}\left[\overline{L}^i(\widehat{p}_i^m, \boldsymbol{\sigma}_{-i}^m) \middle| \overline{L}^i(\widehat{p}_i^m, \boldsymbol{\sigma}_{-i}^m) \geq v_\alpha^i(\widehat{p}_i^m, \boldsymbol{\sigma}_{-i}^m)\right] + \epsilon, \end{aligned} \tag{C.14}$$

where $\widehat{p}_i^m \in \mathit{supp}(CB(\boldsymbol{\sigma}_{-i}^m))$. By adding the inequalities with the same direction in Equations (C.13) and (C.14), for $m \geq \max\{m_5, m_6\}$ we have

$$\begin{aligned} &\mathrm{E}\left[\overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}^m) \middle| \overline{L}^i(p_i', \boldsymbol{\sigma}_{-i}^m) \geq v_\alpha^i(p_i', \boldsymbol{\sigma}_{-i}^m)\right] \\ <&\mathrm{E}\left[\overline{L}^i(\widehat{p}_i^m, \boldsymbol{\sigma}_{-i}^m) \middle| \overline{L}^i(\widehat{p}_i^m, \boldsymbol{\sigma}_{-i}^m) \geq v_\alpha^i(\widehat{p}_i^m, \boldsymbol{\sigma}_{-i}^m)\right] - \epsilon. \end{aligned} \tag{C.15}$$

Equation (C.15) contradicts the fact that $\widehat{p}_i^m \in \mathit{supp}(CB(\boldsymbol{\sigma}_{-i}^m))$, which completes the proof that $\boldsymbol{CB}(\boldsymbol{\sigma})$ has a closed graph.

As listed above, the CVaR_α best response function $\boldsymbol{CB}(\boldsymbol{\sigma})$ satisfies the four conditions of Kakutani's Fixed Point Theorem. As a direct result, for any finite n-player stochastic congestion game, there exists $\boldsymbol{\sigma}^* \in \boldsymbol{\Sigma}$ such that $\boldsymbol{\sigma}^* \in \boldsymbol{CB}(\boldsymbol{\sigma}^*)$, which completes the existence proof of a CVaR_α equilibrium for such games. □

Appendix D

LEMMA PROOFS OF THE BLIND GB-PANDAS ALGORITHM FOR CHAPTER 5

D.1 Proof of Lemma 1

Lemma 1. The following set $\bar{\Lambda}$ is equivalent to Λ defined in equation (5.2):

$$\bar{\Lambda} = \Bigg\{ \boldsymbol{\lambda} = (\lambda_{\bar{L}} : \bar{L} \in \mathcal{L}) \Big| \exists \lambda_{\bar{L},n,m} \geq 0, \forall \bar{L} \in \mathcal{L}, \forall n \in \bar{L}, \forall m \in \mathcal{M}, s.t.$$
$$\lambda_{\bar{L}} = \sum_{n:n\in\bar{L}} \sum_{m=1}^{M} \lambda_{\bar{L},n,m}, \ \forall \bar{L} \in \mathcal{L},$$
$$\sum_{\bar{L}:m\in\bar{L}} \sum_{n:n\in\bar{L}} \frac{\lambda_{\bar{L},n,m}}{\alpha_1} + \sum_{\bar{L}:m\in\bar{L}_2} \sum_{n:n\in\bar{L}} \frac{\lambda_{\bar{L},n,m}}{\alpha_2} +$$
$$\cdots + \sum_{\bar{L}:m\in\bar{L}_N} \sum_{n:n\in\bar{L}} \frac{\lambda_{\bar{L},n,m}}{\alpha_N} < 1, \forall m \Bigg\}, \tag{D.1}$$

where $\lambda_{\bar{L},n,m}$ denotes the arrival rate of type $\bar{L}$ tasks that are 1-local to server n and is processed by server m. $\{\lambda_{\bar{L},n,m} : \bar{L} \in \mathcal{L}, n \in \bar{L}, \text{and } m \in \mathcal{M}\}$ is a decomposition of the set of arrival rates $\{\lambda_{\bar{L},m} : \bar{L} \in \mathcal{L} \text{ and } m \in \mathcal{M}\}$, where $\lambda_{\bar{L},m} = \sum_{n\in\mathcal{M}} \lambda_{\bar{L},n,m}$.

Proof. We show that $\bar{\Lambda} \subset \Lambda$ and $\Lambda \subset \bar{\Lambda}$, which results in the equality of these two sets.

- $\bar{\Lambda} \subset \Lambda$: If $\lambda \in \bar{\Lambda}$, there exists a decomposition $\{\lambda_{\bar{L},n,m} : \bar{L} \in \mathcal{L}, n \in \bar{L}, \text{and } m \in \mathcal{M}\}$ such that the load on each server is less than one under this decomposition. Defining $\lambda_{\bar{L},m} \equiv \sum_{n:n\in\bar{L}} \lambda_{\bar{L},n,m}$, the arrival rate decomposition $\{\lambda_{\bar{L},m} : \bar{L} \in \mathcal{L} \text{ and } m \in \mathcal{M}\}$ obviously satisfies the conditions in the definition of the set Λ, so $\lambda \in \Lambda$ which means that $\bar{\Lambda} \subset \Lambda$.

Portions of this appendix were previously published in Yekkehkhany and Nagi [131] and are used here with permission. Furthermore, portions of this appendix were previously published in Yekkehkhany et al. [200] and are used here with permission.

- $\Lambda \subset \bar{\Lambda}$: If $\lambda \in \Lambda$, there exists a decomposition $\{\lambda_{\bar{L},m} : \bar{L} \in \mathcal{L} \text{ and } m \in \mathcal{M}\}$ such that the load on each server is less than one under this decomposition. Defining $\lambda_{\bar{L},n,m} \equiv \frac{\lambda_{\bar{L},m}}{|\bar{L}|}$, the arrival rate decomposition $\{\lambda_{\bar{L},n,m} : \bar{L} \in \mathcal{L}, n \in \bar{L}, \text{and } m \in \mathcal{M}\}$ obviously satisfies the conditions in the definition of the set $\bar{\Lambda}$, so $\lambda \in \bar{\Lambda}$ which means that $\Lambda \subset \bar{\Lambda}$.

□

D.2 Proof of Lemma 2

Lemma 2.

$$\langle \boldsymbol{W}(t), \widetilde{\boldsymbol{U}}(t) \rangle = 0, \ \forall t.$$

Proof. The expression simplifies as follows:

$$\langle \boldsymbol{W}(t), \widetilde{\boldsymbol{U}}(t) \rangle = \sum_m \left(\frac{Q_m^1(t)}{\alpha_1} + \frac{Q_m^2(t)}{\alpha_2} + \cdots + \frac{Q_m^N(t)}{\alpha_N} \right) \frac{U_m(t)}{\alpha_N}.$$

Note that for any server m, $U_m(t)$ is either zero or positive. For the first case it is obvious that $\left(\frac{Q_m^1(t)}{\alpha_1} + \frac{Q_m^2(t)}{\alpha_2} + \cdots + \frac{Q_m^N(t)}{\alpha_N}\right)\frac{U_m(t)}{\alpha_N} = 0$. In the latter case where $U_m(t) > 0$, all sub-queues of server m are empty which again results in $\left(\frac{Q_m^1(t)}{\alpha_1} + \frac{Q_m^2(t)}{\alpha_2} + \cdots + \frac{Q_m^N(t)}{\alpha_N}\right)\frac{U_m(t)}{\alpha_N} = 0$. Therefore, $\langle \boldsymbol{W}(t), \widetilde{\boldsymbol{U}}(t) \rangle = 0$ for all time slots. □

D.3 Proof of Lemma 3

Lemma 3. Under the GB-PANDAS routing policy, for any arrival rate vector strictly inside the outer bound of the capacity region, $\boldsymbol{\lambda} \in \Lambda$, and the corresponding workload vector of servers $\boldsymbol{w}$ defined in (5.8), we have the following for any t_0:

$$\mathbb{E}\Big[\langle \boldsymbol{W}(t), \boldsymbol{A}(t) \rangle - \langle \boldsymbol{W}(t), \boldsymbol{w} \rangle \Big| Z(t_0) \Big] \leq 0, \ \forall t \geq 0.$$

Proof. The minimum weighted workload for type $\bar{L}$ task, where $\bar{L} \in \mathcal{L}$, at time slot t is defined as follows:

$$W_{\bar{L}}^*(t) = \min_{m \in \mathcal{M}} \left\{ \frac{W_m(t)}{\alpha_1} I_{\{m \in \bar{L}\}}, \frac{W_m(t)}{\alpha_2} I_{\{m \in \bar{L}_2\}}, \cdots, \frac{W_m(t)}{\alpha_N} I_{\{m \in \bar{L}_N\}} \right\}.$$

According to the routing policy of the GB-PANDAS algorithm, an incoming task of type $\bar{L}$ at the beginning of time slot t is routed to the corresponding sub-queue of server m^* with the minimum weighted workload $W^*_{\bar{L}}$. Therefore, for any type $\bar{L}$ task we have the following:

$$\begin{aligned}
&\frac{W_m(t)}{\alpha_1} \geq W^*_{\bar{L}}(t),\ \forall m \in \bar{L},\\
&\frac{W_m(t)}{\alpha_n} \geq W^*_{\bar{L}}(t),\ \forall m \in \bar{L}_n,\quad \text{for } 2 \leq n \leq N.
\end{aligned} \tag{D.2}$$

In other words, a type $\bar{L}$ task does not join a server with a weighted workload greater than $W^*_{\bar{L}}$. Using the fact that $\boldsymbol{W}(t)$ and $\boldsymbol{A}(t)$ are conditionally independent of $Z(t_0)$ given $Z(t)$, and also following the definitions of pseudo task arrival process $\boldsymbol{A}(t)$ in (5.9) and the arrival of an n-local type task to the m-th server $A^n_m(t)$ in (5.4), we have the following:

$$\begin{aligned}
&\mathbb{E}\big[\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle | Z(t_0)\big]\\
&= \mathbb{E}\Big[\mathbb{E}\big[\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle | Z(t)\big]\Big| Z(t_0)\Big]\\
&= \mathbb{E}\Bigg[\mathbb{E}\Bigg[\sum_m W_m(t)\Big(\frac{A^1_m(t)}{\alpha_1} + \frac{A^2_m(t)}{\alpha_2} + \cdots + \frac{A^N_m(t)}{\alpha_N}\Big)\Bigg| Z(t)\Bigg]\Bigg| Z(t_0)\Bigg]\\
&= \mathbb{E}\Bigg[\mathbb{E}\Bigg[\sum_m W_m(t)\Bigg(\frac{1}{\alpha_1}\sum_{\bar{L}:m\in\bar{L}} A_{\bar{L},m}(t) + \frac{1}{\alpha_2}\sum_{\bar{L}:m\in\bar{L}_2} A_{\bar{L},m}(t)\\
&\qquad\qquad + \cdots + \frac{1}{\alpha_N}\sum_{\bar{L}:m\in\bar{L}_N} A_{\bar{L},m}(t)\Bigg)\Bigg| Z(t)\Bigg]\Bigg| Z(t_0)\Bigg]\\
&\overset{(a)}{=} \mathbb{E}\Bigg[\mathbb{E}\Bigg[\sum_{\bar{L}\in\mathcal{L}}\Bigg(\sum_{m:m\in\bar{L}}\frac{W_m(t)}{\alpha_1}A_{\bar{L},m}(t) + \sum_{m:m\in\bar{L}_2}\frac{W_m(t)}{\alpha_2}A_{\bar{L},m}(t) +\\
&\qquad\qquad \cdots + \sum_{m:m\in\bar{L}_N}\frac{W_m(t)}{\alpha_N}A_{\bar{L},m}(t)\Bigg)\Bigg| Z(t)\Bigg]\Bigg| Z(t_0)\Bigg]\\
&\overset{(b)}{=} \mathbb{E}\left[\mathbb{E}\left[\sum_{\bar{L}\in\mathcal{L}} W^*_{\bar{L}}(t)A_{\bar{L}}(t)\Big| Z(t)\right]\Bigg| Z(t_0)\right]\\
&= \sum_{\bar{L}\in\mathcal{L}} W^*_{\bar{L}}(t)\lambda_{\bar{L}},
\end{aligned} \tag{D.3}$$

where (a) is true by changing the order of the summations, and (b) follows by the GB-PANDAS routing policy which routes type $\bar{L}$ task to the server with the minimum weighted workload, $W^*_{\bar{L}}$. Furthermore, using the definition of

the ideal workload on a server in (5.8) we have the following:

$$
\begin{aligned}
&\mathbb{E}\big[\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle | Z(t)\big] \\
&= \sum_{m=1}^{M} W_m(t) w_m \\
&= \sum_{m} W_m(t) \Bigg(\sum_{\bar{L}: m\in\bar{L}} \frac{\lambda_{\bar{L},m}}{\alpha_1} + \sum_{\bar{L}: m\in\bar{L}_2} \frac{\lambda_{\bar{L},m}}{\alpha_2} + \cdots + \sum_{\bar{L}: m\in\bar{L}_N} \frac{\lambda_{\bar{L},m}}{\alpha_N} \Bigg) \\
&\overset{(a)}{=} \sum_{\bar{L}\in\mathcal{L}} \Bigg(\sum_{m: m\in\bar{L}} \frac{W_m(t)}{\alpha_1} \lambda_{\bar{L},m} + \sum_{m: m\in\bar{L}_2} \frac{W_m(t)}{\alpha_2} \lambda_{\bar{L},m} + \\
&\qquad\qquad \cdots + \sum_{m: m\in\bar{L}_N} \frac{W_m(t)}{\alpha_N} \lambda_{\bar{L},m} \Bigg) \\
&\overset{(b)}{\geq} \sum_{\bar{L}\in\mathcal{L}} \sum_{m\in\mathcal{M}} W^*_{\bar{L}}(t) \lambda_{\bar{L},m} \\
&= \sum_{\bar{L}\in\mathcal{L}} W^*_{\bar{L}}(t) \lambda_{\bar{L}},
\end{aligned}
\tag{D.4}
$$

where (a) is true by changing the order of summations, and (b) follows from (D.2). Lemma 3 is concluded from equations (D.3) and (D.4). □

D.4 Proof of Lemma 4

Lemma 4. Under the GB-PANDAS routing policy, for any arrival rate vector strictly inside the outer bound of the capacity region, $\boldsymbol{\lambda} \in \Lambda$, and the corresponding workload vector of servers $\boldsymbol{w}$ defined in (5.8) there exists $T_0 > 0$ such that for any $T \geq T_0$ we have the following:

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1} \Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle \Big) \Big| Z(t_0) \right] \\
&\leq -\theta_0 T ||\boldsymbol{Q}(t_0)||_1 + c_0, \ \forall t_0 \geq 0,
\end{aligned}
$$

where the constants $\theta_0, c_0 > 0$ are independent of $Z(t_0)$.

Proof. By our assumption on boundedness of arrival and service processes, there exists a constant C_A such that for any t_0, t, and T with $t_0 \leq t \leq t_0 + T$,

we have the following:

$$W_m(t_0) - \frac{T}{\alpha_N} \leq W_m(t) \leq W_m(t_0) + \frac{TC_A}{\alpha_N},\ \forall m \in \mathcal{M}. \tag{D.5}$$

On the other hand, by (5.7) the ideal workload on a server defined in (5.8) can be bounded as follows:

$$w_m \leq \frac{1}{1+\delta},\ \forall m \in \mathcal{M}. \tag{D.6}$$

Hence,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|Z(t_0)\right] \\
&= \mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\sum_{m=1}^{M} W_m(t)w_m\right)\Big|Z(t_0)\right] \\
&\overset{(a)}{\leq} T\sum_{m=1}^{M}\Big(W_m(t_0)w_m\Big) + \frac{MT^2C_A}{\alpha_N} \\
&\overset{(b)}{\leq} \frac{T}{1+\delta}\sum_{m} W_m(t_0) + \frac{MT^2C_A}{\alpha_N},
\end{aligned} \tag{D.7}$$

where (a) is true by bringing the inner summation on m out of the expectation and using the boundedness property of the workload in equation (D.5), and (b) is true by Equation (D.6).

Before investigating the second term, $\mathbb{E}\Big[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|Z(t_0)\Big]$, we propose the following lemma which will be used in lower bounding this second term.

Lemma 13. *For any server $m \in \mathcal{M}$ and any t_0, we have the following:*

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_1} + \frac{S_m^2(t)}{\alpha_2} + \cdots + \frac{S_m^N(t)}{\alpha_N}\right)\Big|Z(t_0)\right]}{T} = 1.$$

The proof of Lemma 13 is provided in Appendix D.5. We then have the following:

$$\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|Z(t_0)\right]$$

$$
\begin{aligned}
&= \mathbb{E}\Bigg[\sum_{t=t_0}^{t_0+T-1}\sum_{m=1}^{M}\left(W_m(t)\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\right)\Bigg|Z(t_0)\Bigg]\\
&\overset{(a)}{\geq} \sum_{m=1}^{M}\left(W_m(t_0)\mathbb{E}\Bigg[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\Bigg|Z(t_0)\Bigg]\right)\\
&\quad-\frac{T}{\alpha_N}\sum_{m=1}^{M}\mathbb{E}\Bigg[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\Bigg|Z(t_0)\Bigg],
\end{aligned}
\tag{D.8}
$$

where (a) follows by bringing the inner summation on m out of the expectation and using the boundedness property of the workload in equation (D.5).

Using Lemma 13, for any $0 < \epsilon_0 < \frac{\delta}{1+\delta}$, there exists T_0 such that for any $T \geq T_0$, we have the following for any server $m \in \mathcal{M}$:

$$
1-\epsilon_0 \leq \frac{\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\Big|Z(t_0)\right]}{T} \leq 1+\epsilon_0.
$$

Then continuing on equation (D.8) we have the following:

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|Z(t_0)\right]\\
&\geq T(1-\epsilon_0)\sum_{m=1}^{M}W_m(t_0)-\frac{MT^2(1+\epsilon_0)}{\alpha_N}.
\end{aligned}
\tag{D.9}
$$

Then, Lemma 4 is concluded as follows by using equations (D.7) and (D.9) and picking $c_0 = \frac{MT^2}{\alpha_N}(C_A+1+\epsilon_0)$ and $\theta_0 = \frac{1}{\alpha_1}\left(\frac{\delta}{1+\delta}-\epsilon_0\right)$, where by our choice of ϵ_0 we have $\theta_0 > 0$:

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|Z(t_0)\right]\\
&\leq -T\left(\frac{\delta}{1+\delta}-\epsilon_0\right)\sum_{m=1}^{M}W_m(t_0)+\frac{MT^2}{\alpha_N}(C_A+1+\epsilon_0)\\
&\overset{(a)}{\leq} -\frac{T}{\alpha_1}\left(\frac{\delta}{1+\delta}-\epsilon_0\right)\sum_{m=1}^{M}\Big(Q_m^1(t_0)+Q_m^2(t_0)+\cdots+Q_m^N(t_0)\Big)+c_0\\
&\leq -\theta_0 T||\boldsymbol{Q}(t_0)||_1+c_0,\ \forall t_0 \geq 0,
\end{aligned}
$$

where (a) is true as $W_m(t_0) \geq \frac{Q_m^1(t_0)+Q_m^2(t_0)+\cdots+Q_m^N(t_0)}{\alpha_1}$. □

D.5 Proof of Lemma 13

Lemma 13. For any server $m \in \mathcal{M}$ and any t_0, we have the following:

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\middle| Z(t_0)\right]}{T} = 1.$$

Proof. Let t_m^* be the first time slot after or at time slot t_0 at which server m becomes idle, and so is available to serve another task; that is,

$$t_m^* = \min\{\tau : \tau \geq t_0, \Psi_m(\tau) = 0\}, \tag{D.10}$$

where, as a reminder, $\Psi_m(\tau)$ is the number of time slots that the m-th server has spent on the task that is receiving service from this server at time slot τ. Note that the CDF of the service time distributions are given by $F_n, n \in \{1, 2, \cdots, N\}$ where they all have finite means $\alpha_n < \infty$; therefore, $t_m^* < \infty$. We then have the following by considering the bounded service:

$$\begin{aligned}
&\left(\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1}\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\middle| Z(t_0)\right] - \frac{t_m^*-t_0}{\alpha_N}+\frac{1}{\alpha_1}\right)\Big/T \\
&\leq \frac{\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\middle| Z(t_0)\right]}{T} \leq \\
&\left(\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1}\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\middle| Z(t_0)\right] + \frac{1}{\alpha_N}\right)\Big/T,
\end{aligned} \tag{D.11}$$

where by boundedness of t_m^*, α_1, and α_N, it is obvious that $\lim_{T\to\infty} \frac{-\frac{t_m^*-t_0}{\alpha_N}+\frac{1}{\alpha_1}}{T} = 0$ and $\lim_{T\to\infty} \frac{\frac{1}{\alpha_N}}{T} = 0$. Hence, by taking the limit of the terms in equation (D.11) as T goes to infinity, we have the following:

$$\begin{aligned}
&\lim_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\middle| Z(t_0)\right]}{T} \\
&= \lim_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1}\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\middle| Z(t_0)\right]}{T}.
\end{aligned} \tag{D.12}$$

Considering the service process as a renewal process, given the scheduling decisions at the end of the renewal intervals in $[t_m^*, t_m^* + T - 1]$, all holding times for server m to give service to tasks in its queues are independent. We elaborate on this in the following.

We define renewal processes, $N_m^n(t),\ n \in \{1, 2, \cdots, N\}$, as follows, where t is an integer valued number: Let $H_m^n(l)$ be the holding time (service time) of the l-th task that is n-local to server m after time slot t_m^* receiving service from serve m, and call $\{H_m^n(l), l \geq 1\}$ the holding process of n-local type task ($n \in \{1, 2, \cdots, N\}$). Then define $J_m^n(l) = \sum_{i=1}^{l} H_m^n(l)$ for $l \geq 1$, and let $J_m^n(0) = 0$. In the renewal process, $J_m^n(l)$ is the l-th jumping time, or the time at which the l-th occurrence happens, and it has the following relation with the renewal process, $N_m^n(t)$:

$$N_m^n(t) = \sum_{l=1}^{\infty} \mathbb{I}_{\{J_m^n(l) \leq t\}} = \sup\{l : J_m^n(l) \leq t\}.$$

Another way to define $N_m^n(t)$ is as below:

Set $\tau = t_m^*$, $cntr = 0$, $N_m^n(t) = 0$
while $cntr < t$ **do**
  if $\eta_m(\tau) = n$ **then**
    $cntr++$
    $N_m^n(t) += S_m^n(\tau)$
  end if
  $\tau++$
end while

By convention, $N_m^n(0) = 0$.

In the following, we define another renewal process, $N_m(t)$:

$$N_m(t) = \sum_{u=t_m^*}^{t_m^*+t-1} \left(\mathbb{I}_{\{S_m^1(u)=1\}} + \mathbb{I}_{\{S_m^2(u)=1\}} + \cdots + \mathbb{I}_{\{S_m^N(u)=1\}} \right).$$

Similarly, let $H_m(l)$ be the holding time (service time) of the l-th task after time slot t_m^* receiving service from serve m, and call $\{H_m(l), l \geq 1\}$ the holding process. Then define $J_m(l) = \sum_{i=1}^{l} H_m(l)$ for $l \geq 1$, and let $J_m(0) = 0$. In the renewal process, $J_m(l)$ is the l-th jumping time, or the time at which the l-th occurrence happens, and it has the following relation

with the renewal process, $N_m(t)$:

$$N_m(t) = \sum_{l=1}^{\infty} \mathbb{I}_{\{J_m(l) \le t\}} = \sup\{l : J_m(l) \le t\}.$$

Note that the central scheduler makes scheduling decisions for server m at time slots $\{t_m^* + J_m(l), l \ge 1\}$. We denote these scheduling decisions by $D_m(t_m^*) = \Big(\eta_m(t_m^* + J_m(l)) : l \ge 1\Big)$.

Consider the time interval $[t_m^*, t_m^* + T - 1]$ when T goes to infinity. Define ρ_m^n as the fraction of time that server m is busy giving service to tasks that are n-local to this server, in the mentioned interval. Obviously, $\sum_{n=1}^{N} \rho_m^n = 1$. Then equation (D.12) is followed by the following:

$$\begin{aligned}
&\lim_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1} \left(\frac{S_m^1(t)}{\alpha_1} + \frac{S_m^2(t)}{\alpha_2} + \cdots + \frac{S_m^N(t)}{\alpha_N}\right)\Big| Z(t_0)\right]}{T} \\
&= \lim_{T\to\infty} \Bigg\{ \mathbb{E}\Bigg[\mathbb{E}\Bigg[\sum_{t=t_m^*}^{t_m^*+T-1} \Bigg(\frac{S_m^1(t)}{\alpha_1} + \frac{S_m^2(t)}{\alpha_2} + \\
&\qquad\qquad \cdots + \frac{S_m^N(t)}{\alpha_N}\Bigg)\Big| D_m(t_m^*), Z(t_0)\Bigg]\Big| Z(t_0)\Bigg]\Bigg\} \Big/ T \\
&= \sum_{n=1}^{N} \lim_{T\to\infty} \left(\mathbb{E}\left[\frac{1}{\alpha_n}\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1} \Big(S_m^n(t)\Big)\Big| D_m(t_m^*), Z(t_0)\right]\Big| Z(t_0)\right]\right) \Big/ T \\
&= \sum_{n=1}^{N} \mathbb{E}\left[\frac{1}{\alpha_n} \lim_{T\to\infty} \frac{\mathbb{E}\left[N_m^n\big(\rho_m^n T\big)\Big| D_m(t_m^*), Z(t_0)\right]}{T}\Big| Z(t_0)\right].
\end{aligned} \tag{D.13}$$

Note that given $\{D_m(t_m^*), Z(t_0)\}$, the holding times $\{H_m^n(l), l \ge 1\}$ are independent and identically distributed with CDF F_n. If $\rho_m^n = 0$, then we do not have to worry about those tasks that are n-local to server m since they receive service from this server for only a finite number of times in time interval $[t_m^*, t_m^* + T - 1]$ as $T \to \infty$, so

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[N_m^n\big(\rho_m^n T\big)\big| D_m(t_m^*), Z(t_0)\right]}{T} = 0.$$

But if $\rho_m^n > 0$, we can use the strong law of large numbers for renewal process

N_m^n to conclude the following:

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[N_m^n\left(\rho_m^n T\right)\middle| D_m(t_m^*), Z(t_0)\right]}{T} = \rho_m^n \cdot \frac{1}{\mathbb{E}[H_m^n(1)]}, \tag{D.14}$$

where the holding time (service time) $H_m^n(1)$ has CDF F_n with expectation $\frac{1}{\alpha_n}$. Combining equations (D.15) and (D.14), Lemma 13 is concluded as follows:

$$\begin{aligned}&\lim_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1}\left(\frac{S_m^1(t)}{\alpha_1}+\frac{S_m^2(t)}{\alpha_2}+\cdots+\frac{S_m^N(t)}{\alpha_N}\right)\middle| Z(t_0)\right]}{T}\\ =&\sum_{n=1}^{N}\mathbb{E}\left[\frac{1}{\alpha_n}\cdot\rho_m^n\cdot\alpha_n\middle| Z(t_0)\right] = \sum_{n=1}^{N}\rho_m^n = 1.\end{aligned} \tag{D.15}$$

□

D.6 Proof of Lemma 5

Lemma 5. Under the GB-PANDAS routing policy, for any arrival rate vector strictly inside the outer bound of the capacity region, $\boldsymbol{\lambda} \in \Lambda$, and any $\theta_1 \in (0,1)$, there exists $T_1 > 0$ such that the following is true for any $T \geq T_1$ and for any $t_0 \geq 0$:

$$\begin{aligned}&\mathbb{E}\Big[||\boldsymbol{\Psi}(t_0+T)||_1 - ||\boldsymbol{\Psi}(t_0)||_1\Big| Z(t_0)\Big]\\ \leq& -\theta_1||\boldsymbol{\Psi}(t_0)||_1 + MT,\end{aligned}$$

where $||.||_1$ is L^1-norm.

Proof. For any server $m \in \mathcal{M}$, let t_m^* be the first time slot after or at time slot t_0 at which the server is available (t_m^* is also defined in (D.10)); that is,

$$t_m^* = \min\{\tau : \tau \geq t_0, \Psi_m(\tau) = 0\}, \tag{D.16}$$

where it is obvious that $\Psi_m(t_m^*) = 0$.
Note that for any t, we have $\Psi_m(t+1) \leq \Psi_m(t) + 1$, that is true by the definition of $\Psi(t)$, which is the number of time slots that server m has spent on

the currently in-service task. From time slot t to $t+1$, if a new task comes in service, then $\Psi_m(t+1) = 0$ which results in $\Psi_m(t+1) \leq \Psi_m(t)+1$; otherwise, if server m continues giving service to the same task, then $\Psi_m(t+1) = \Psi_m(t)+1$. Thus, if $t_m^* \leq t_0+T$, it is easy to find out that $\Psi_m(t_0+T) \leq t_0+T-t_m^* \leq T$. In the following we use t_m^* to find a bound on $\mathbb{E}[\Psi_m(t_0+T) - \Psi_m(t_0)|Z(t_0)]$:

$$
\begin{aligned}
&\mathbb{E}\Big[||\boldsymbol{\Psi}(t_0+T)||_1 - ||\boldsymbol{\Psi}(t_0)||_1 \Big| Z(t_0)\Big] \\
&= \sum_{m=1}^{M} \mathbb{E}\left[\Big(\Psi_m(t_0+T) - \Psi_m(t_0)\Big)\Big| Z(t_0)\right] \\
&= \sum_{m=1}^{M} \Bigg\{ \mathbb{E}\left[\Big(\Psi_m(t_0+T) - \Psi_m(t_0)\Big)\Big| Z(t_0), t_m^* \leq t_0+T\right] \\
&\qquad\qquad \times P\left(t_m^* \leq t_0+T \big| Z(t_0)\right) \\
&\qquad + \mathbb{E}\left[\Big(\Psi_m(t_0+T) - \Psi_m(t_0)\Big)\Big| Z(t_0), t_m^* > t_0+T\right] \\
&\qquad\qquad \times P\left(t_m^* > t_0+T \big| Z(t_0)\right) \Bigg\} \qquad \text{(D.17)} \\
&\overset{(a)}{\leq} \sum_{m=1}^{M} \Bigg\{ \Big(T - \Psi_m(t_0)\Big) \times P\left(t_m^* > t_0+T \big| Z(t_0)\right) \\
&\qquad\qquad + T \times P\left(t_m^* > t_0+T \big| Z(t_0)\right) \Bigg\} \\
&= -\sum_{m=1}^{M} \Big(\Psi_m(t_0) \cdot P\left(t_m^* > t_0+T \big| Z(t_0)\right)\Big) + MT,
\end{aligned}
$$

where (a) is true as given that $t_m^* \leq t_0+T$ we found that $\Psi_m(t_0+T) \leq T$, so $\Psi_m(t_0+T) - \Psi_m(t_0) \leq T - \Psi_m(t_0)$, and given that $t_m^* > t_0+T$, it is concluded that server m is giving service to the same task over the whole interval $[t_0, t_0+T]$, which results in $\Psi_m(t_0+T) - \Psi_m(t_0) = T$.
Since service time of an n-local task has CDF F_n with finite mean, we have the following:

$$\lim_{T\to\infty} P\left(t_m^* \leq t_0+T \Big| Z(t_0)\right) = 1, \ \forall m \in \mathcal{M}.$$

Therefore, for any $\theta_1 \in (0,1)$ there exists T_1 such that for any $T \geq T_1$, we have $P\left(t_m^* \leq t_0+T \Big| Z(t_0)\right) \geq \theta_1$, for any $m \in \mathcal{M}$, so equation (D.17) follows as below which completes the proof:

$$
\begin{aligned}
&\mathbb{E}\Big[||\boldsymbol{\Psi}(t_0+T)||_1 - ||\boldsymbol{\Psi}(t_0)||_1 \Big| Z(t_0)\Big] \\
\leq &-\theta_1 \sum_{m=1}^{M} \Psi_m(t_0) + MT \\
= &-\theta_1 ||\boldsymbol{\Psi}(t_0)||_1 + MT.
\end{aligned} \tag{D.18}
$$

□

D.7 Proof of Lemma 6

Lemma 6. $\big\{\boldsymbol{Z}(t) = \big(\boldsymbol{Q}(t), \boldsymbol{\eta}(t), \boldsymbol{\Psi}(t)\big), t \geq 0\big\}$ forms an irreducible and aperiodic Markov chain. The state space of this Markov chain is $\mathcal{S} = \big(\prod_{m\in\mathcal{M}} \mathbb{N}^{N^m}\big) \times \big(\prod_{m\in\mathcal{M}} \{1, 2, \cdots, N^m\}\big) \times \mathbb{N}^M$.

Proof. Consider $\boldsymbol{Z}(0) = \big\{0_{(\sum_{m\in\mathcal{M}} N^m)\times 1}, \prod_{m\in\mathcal{M}} N^m,\ 0_{M\times 1}\big\}$ as the initial state of the Markov chain $\boldsymbol{Z}(t)$.

Irreducible: Since $F_{i,m}$ is increasing for any task-server pair, we can find an integer $\tau > 0$ such that $F_{i,m}(\tau) > 0$ for any $1 \leq i \leq N^m$ and $m \in \mathcal{M}$. Furthermore, probability of zero task arrival is positive in each time slot. Hence, for any state $\boldsymbol{Z} = (\boldsymbol{Q}, \boldsymbol{\eta}, \boldsymbol{\Psi})$, there is a positive probability that each task receives service in τ time slots and no new task arrives at the system in $\tau \sum_{m\in\mathcal{M}} \sum_{n=1}^{N^m} Q_m^n$ time slots. Accordingly, the initial state of the Markov chain is reachable from any states of the system. Conversely, using the same approach, it is easy to see that any states of the system is reachable from the initial state, $\boldsymbol{Z}(0)$. Consequently, the Markov chain $\boldsymbol{Z}(t)$ is irreducible.

Aperiodic: Since Markov chain $\boldsymbol{Z}(t)$ is irreducible, in order to show that it is also aperiodic, it suffices to show that there is a positive probability for transition from a state to itself. Due to the fact that there is a positive probability that zero task arrives to the system, the Markov chain stays at the initial state with a positive probability. Hence, the Markov chain $\boldsymbol{Z}(t)$ is aperiodic. □

D.8 Proof of Lemma 7

Lemma 7. For any arrival rate vector inside the capacity region, $\boldsymbol{\lambda} \in \Lambda$, there exists a load decomposition $\{\lambda_{i,m}\}$ and $\delta > 0$ such that

$$\sum_{i \in \mathcal{L}} \frac{\lambda_{i,m}}{\mu_{i,m}} < \frac{1}{1+\delta}, \ \forall m \in \mathcal{M}. \tag{D.19}$$

The fluid model planning algorithm solves a linear programming to find the load decomposition $\{\lambda_{i,m}\}$ that is used in its load balancing on the M servers. In other words, this load decomposition is a possibility of task assignment on servers to stabilize the system.

Proof. The capacity region Λ is an open set, so for any $\boldsymbol{\lambda} \in \Lambda$, there exists $\delta > 0$ such that $(1+\delta)\boldsymbol{\lambda} = \boldsymbol{\lambda}' \in \Lambda$. On that account, (5.15) follows by $\sum_{i \in \mathcal{L}} \frac{\lambda'_{i,m}}{\mu_{i,m}} = \sum_{i \in \mathcal{L}} \frac{(1+\delta)\lambda_{i,m}}{\mu_{i,m}} < 1, \forall m \in \mathcal{M}$, which completes the proof:

$$\sum_{i \in \mathcal{L}} \frac{\lambda_{i,m}}{\mu_{i,m}} < \frac{1}{1+\delta}, \forall m \in \mathcal{M}.$$

□

D.9 Proof of Lemma 8

Lemma 8.

$$\langle \boldsymbol{W}(t), \widetilde{\boldsymbol{U}}(t) \rangle = 0, \ \forall t.$$

Proof.

$$\langle \boldsymbol{W}(t), \widetilde{\boldsymbol{U}}(t) \rangle = \sum_{m \in \mathcal{M}} \left(\frac{Q_m^1(t)}{\alpha_m^1} + \frac{Q_m^2(t)}{\alpha_m^2} + \cdots + \frac{Q_m^N(t)}{\alpha_m^{N^m}} \right) \frac{U_m(t)}{\alpha_m^{N^m}}.$$

If the unused service for server m is zero, $U_m(t) = 0$, the corresponding term for server m is zero in the above summation. Alternatively, the unused service of server m is positive if and only if all N^m sub-queues of the server are empty, which again makes the corresponding term for server m in the above summation equal to zero. □

D.10 Proof of Lemma 9

Lemma 9. Under the exploration-exploitation routing policy of the Blind GB-PANDAS algorithm, for any arrival rate vector inside the capacity region, $\boldsymbol{\lambda} \in \Lambda$, and the corresponding ideal workload vector $\boldsymbol{w}$ defined in (5.24), and for any arbitrary small $\theta_0 > 0$, there exists $T_0 > t_0$ such that for any $t_0 \geq 0$ and $T > T_0$:

$$\begin{aligned}&\mathbb{E}\Big[\sum_{t=T_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|\boldsymbol{Z}(t_0)\Big]\\ \leq&\theta_0 T\|\boldsymbol{Q}(t_0)\|_1 + c_0,\end{aligned}$$

where the constants $\theta_0, c_0 > 0$ are independent of $\boldsymbol{Z}(t_0)$.

Proof. By the choice of exploration rate for Blind GB-PANDAS, which is independent of the system state, and the fact that exploration exists in both routing and scheduling, any task that is n-local to server m is scheduled on this server for infinitely many times in the interval $[t_0, \infty)$ only due to exploration, regardless of the initial system state. Processing time of an n-local task on server m has a finite mean. Hence, due to strong law of large numbers, using the update rule (5.17) for the elements of the service rate matrix, we have:

$$\begin{aligned}&\forall\ 0 < \epsilon < \frac{1}{2} \times \min\Big\{\min_{n\neq n',m}\big|\alpha_m^n - \alpha_m^{n'}\big|, \min_{m,n}\alpha_m^n, 0.5\Big\}\\ &\text{and } \forall \delta' > 0, \exists T_0' > t_0, \text{ such that for any } \boldsymbol{Z}(t_0)\\ &P\bigg(\big|\widetilde{\alpha}_m^n(t) - \alpha_m^n\big| < \epsilon,\ 1-\epsilon < \frac{\alpha_m^n}{\widetilde{\alpha}_m^n(t)} < 1+\epsilon\bigg|\boldsymbol{Z}(t_0)\bigg) > 1-\delta',\\ &\forall t > T_0',\ \forall m \in \mathcal{M},\ \forall n \in \{1, 2, \cdots, N^m\}.\end{aligned} \tag{D.20}$$

By the above choice of ϵ, for $t > T_0'$, the different locality levels are distinct from each other with at least $1-\delta'$ probability. Let E be the event that $\big|\widetilde{\alpha}_m^n(t) - \alpha_m^n\big| < \epsilon$ and $1-\epsilon < \frac{\alpha_m^n}{\widetilde{\alpha}_m^n(t)} < 1+\epsilon$ for $t \geq T_0'$.

For an incoming task of type $i \in \mathcal{L}$ at time slot t, define the *exact* (but

not known) and *estimated* minimum weighted workloads as

$$\begin{aligned}\overline{W}_i^*(t) &= \min_{m\in\mathcal{M}} \frac{W_m(t)}{\mu_{i,m}},\\ \overline{\widetilde{W}}_i^*(t) &= \min_{m\in\mathcal{M}} \frac{\widetilde{W}_m(t)}{\widetilde{\mu}_{i,m}(t)},\end{aligned} \tag{D.21}$$

where $W_m(t)$ and $\widetilde{W}_m(t)$ are defined in (5.13) and (5.18), respectively. $W_m(t)$ and $\widetilde{W}_m(t)$ are related to each other as follows:

$$\begin{aligned}\widetilde{W}_m(t) &= \frac{Q_m^1(t)}{\widetilde{\alpha}_m^1(t)} + \frac{Q_m^2(t)}{\widetilde{\alpha}_m^2(t)} + \cdots + \frac{Q_m^{N^m}(t)}{\widetilde{\alpha}_m^{N^m}(t)}\\ &= \frac{\alpha_m^1}{\widetilde{\alpha}_m^1(t)}\cdot\frac{Q_m^1(t)}{\alpha_m^1} + \cdots + \frac{\alpha_m^{N^m}}{\widetilde{\alpha}_m^{N^m}(t)}\cdot\frac{Q_m^{N^m}(t)}{\alpha_m^{N^m}}.\end{aligned}$$

Hence, using (D.20), for any $t > T_0'$ and any $m \in \mathcal{M}$, we have

$$P\left((1-\epsilon)W_m(t) < \widetilde{W}_m(t) < (1+\epsilon)W_m(t)\middle|\boldsymbol{Z}(t_0), E\right) = 1, \tag{D.22}$$

and using (D.21) and (D.22), we have

$$P\left(\frac{W_m(t)}{\mu_{i,m}} \geq \overline{W}_i^*(t) > \frac{1}{(1+\epsilon)^2}\overline{\widetilde{W}}_i^*(t)\middle|\boldsymbol{Z}(t_0), E\right) = 1. \tag{D.23}$$

Using the conditional independence of $\widetilde{\boldsymbol{W}}(t)$ and $\boldsymbol{A}(t)$ from $\boldsymbol{Z}(t_0)$ given $\boldsymbol{Z}(t)$, for any $T > T_0' - t_0$, we have the following for $T_0' \leq t \leq t_0 + T - 1$:

$$\begin{aligned}&\mathbb{E}\big[\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle | \boldsymbol{Z}(t_0)\big]\\ &\overset{(a)}{=} \mathbb{E}\Big[\sum_{m\in\mathcal{M}} W_m(t)\Big(\frac{A_m^1(t)}{\alpha_m^1} + \frac{A_m^2(t)}{\alpha_m^2} + \cdots + \frac{A_m^{N^m}(t)}{\alpha_m^{N^m}}\Big)\Big|\boldsymbol{Z}(t_0)\Big]\\ &\overset{(b)}{=} \mathbb{E}\Big[\sum_m W_m(t)\Big(\frac{1}{\alpha_m^1}\sum_{i\in\mathcal{L}_m^1} A_{i,m}(t) + \frac{1}{\alpha_m^2}\sum_{i\in\mathcal{L}_m^2} A_{i,m}(t)\\ &\qquad + \cdots + \frac{1}{\alpha_m^{N^m}}\sum_{i\in\mathcal{L}_m^{N^m}} A_{i,m}(t)\Big)\Big|\boldsymbol{Z}(t_0)\Big]\\ &\overset{(c)}{=} \mathbb{E}\Big[\sum_{i\in\mathcal{L}}\sum_{m\in\mathcal{M}}\Big(\frac{W_m(t)}{\mu_{i,m}}A_{i,m}(t)\Big)\Big|\boldsymbol{Z}(t_0)\Big]\end{aligned}$$

$$
\begin{aligned}
&\overset{(d)}{\leq} \mathbb{E}\Bigg[\sum_{i\in\mathcal{L}}\sum_{m\in\mathcal{M}}\left(\frac{1}{(1-\epsilon)^2}\cdot\frac{\widetilde{W}_m(t)}{\widetilde{\mu}_{i,m}}A_{i,m}(t)\right)\Bigg|\boldsymbol{Z}(t_0),E\Bigg]\\
&\quad+\delta'\cdot\mathbb{E}\Bigg[\sum_{i\in\mathcal{L}}\sum_{m\in\mathcal{M}}\left(\frac{Q_m^1(t)+\cdots+Q_m^{N^m}(t)}{\min_{i,m}\{\mu_{i,m}\}\cdot\min_i\{\mu_{i,m}\}}A_{i,m}(t)\right)\Bigg|\boldsymbol{Z}(t_0),E^c\Bigg]\\
&\overset{(e)}{<} \mathbb{E}\Bigg[\mathbb{E}\Bigg[\sum_{i\in\mathcal{L}}\Bigg(p_e\cdot\frac{1}{(1-\epsilon)^2}\cdot\widetilde{\overline{W}}_i^*(t)A_i(t)+\frac{1-p_e}{(1-\epsilon)^2}\\
&\quad\times\sum_m\frac{\sum_{n=1}^{N^m}Q_m^n(t_0)+N_T(T-t_0)C_A}{\min_{i,m}\{\widetilde{\mu}_{i,m}(t)\}\cdot\min_i\{\widetilde{\mu}_{i,m}(t)\}}\cdot C_A\Bigg)\Bigg|\boldsymbol{Z}(t)\Bigg]\Bigg|\boldsymbol{Z}(t_0)\,E\Bigg]\\
&\quad+\delta'\cdot\mathbb{E}\Bigg[\sum_{i\in\mathcal{L}}\Bigg(\sum_m\frac{\sum_{n=1}^{N^m}Q_m^n(t_0)+N_T(T-t_0)C_A}{\min_{i,m}\{\mu_{i,m}\}\cdot\min_i\{\mu_{i,m}\}}\cdot C_A\Bigg)\Bigg|\boldsymbol{Z}(t_0),E^c\Bigg]\\
&\overset{(f)}{<} \frac{1}{(1-\epsilon)^2}\sum_{i\in\mathcal{L}}\mathbb{E}\Bigg[\widetilde{\overline{W}}_i^*(t)\Bigg|\boldsymbol{Z}(t_0),E\Bigg]\lambda_i+\left(\frac{1}{t^{\delta''}}+\delta'\right)c_0''\|\boldsymbol{Q}(t_0)\|_1+c_0',
\end{aligned}
\tag{D.24}
$$

where (a) and (b) are simply followed by the definitions of pseudo task arrival process in (5.21) and $A_m^n(t)$ in (5.19), respectively. The order of summations is changed in (c). By the law of total probability, (D.20), and (D.22), (d) is true, and (e) follows by the routing policy of Blind GB-PANDAS, where an incoming task at the beginning of time slot t is routed to the corresponding sub-queue of the server with the minimum estimated weighted workload with probability $p_e = \max(1-p(t),0)$ and is routed to the corresponding sub-queue of a server chosen uniformly at random with probability $1-p_e$. Also note that the number of arriving tasks at a time slot is assumed to be upper bounded by C_A. The last step, (f), is true by using (D.20), upper bounding the exploration probability $1-p_e$ by $\frac{1}{t^{\delta''}}$ given that $\delta''>0$ is a constant, and doing simple calculations, where c_0'' and c_0' are constants independent of $\boldsymbol{Z}(t_0)$. Note that minimum value of the estimated service rates, $\min_{i,m}\{\widetilde{\mu}_{i,m}(t)\}$, is lower bounded for any $t\geq t_0$ by a constant which is the minimum of the initialization of service rates and the inverse of the maximum support of CDF functions $F_{i,m}$. We also have

$$
\begin{aligned}
&\mathbb{E}\big[\langle\boldsymbol{W}(t),\boldsymbol{w}\rangle|\boldsymbol{Z}(t_0)\big]\\
&=\mathbb{E}\left[\sum_{m\in\mathcal{M}}W_m(t)w_m\Big|\boldsymbol{Z}(t_0)\right]\\
&\overset{(a)}{=}\mathbb{E}\left[\sum_{m\in\mathcal{M}}\Big(W_m(t)\sum_{i\in\mathcal{L}}\frac{\lambda_{i,m}}{\mu_{i,m}}\Big)\Big|\boldsymbol{Z}(t_0)\right]
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(b)}{=} \mathbb{E}\bigg[\sum_{\substack{i\in\mathcal{L}\\ m\in\mathcal{M}}} \frac{W_m(t)}{\mu_{i,m}}\lambda_{i,m}\bigg|\boldsymbol{Z}(t_0)\bigg] \\
&\overset{(c)}{\geq} \sum_{\substack{i\in\mathcal{L}\\ m\in\mathcal{M}}} \frac{1-\delta'}{(1+\epsilon)^2}\mathbb{E}\bigg[\widetilde{\overline{W}}_i^*(t)\bigg|\boldsymbol{Z}(t_0),E\bigg]\lambda_{i,m} \qquad \text{(D.25)}\\
&= \frac{1-\delta'}{(1+\epsilon)^2}\sum_{i\in\mathcal{L}}\mathbb{E}\bigg[\widetilde{\overline{W}}_i^*(t)\bigg|\boldsymbol{Z}(t_0),E\bigg]\lambda_i,
\end{aligned}
$$

where (a) is true by the definition of the ideal workload on a server in (5.24), note that the ideal workload is not state dependent but $W_m(t)$ is, the order of summations is changed in (b), and (c) is followed by the law of total probability, ignoring the second term, and Equation (D.23).

Putting (D.24) and (D.25) together, for $T > T_0 > T_0'$, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\sum_{t=T_0}^{t_0+T-1}\Big(\langle\boldsymbol{W}(t),\boldsymbol{A}(t)\rangle - \langle\boldsymbol{W}(t),\boldsymbol{w}\rangle\Big)\Big|\boldsymbol{Z}(t_0)\Big] \\
&< \sum_{t=T_0}^{t_0+T-1}\Bigg(\bigg(\frac{1}{(1-\epsilon)^2}-\frac{1-\delta'}{(1+\epsilon)^2}\bigg)\sum_{i\in\mathcal{L}}\mathbb{E}\bigg[\widetilde{\overline{W}}_i^*(t)\bigg|\boldsymbol{Z}(t_0),E\bigg]\lambda_i \\
&\quad + \bigg(\frac{1}{t^{\delta''}}+\delta'\bigg)c_0''\|\boldsymbol{Q}(t_0)\|_1 + c_0'\Bigg) \\
&\overset{(a)}{<} \frac{16}{9}(4\epsilon+\delta')\cdot\Bigg(\sum_{t=T_0}^{t_0+T-1}\sum_{i\in\mathcal{L}}\mathbb{E}\bigg[\widetilde{\overline{W}}_i^*(t)\bigg|\boldsymbol{Z}(t_0),E\bigg]\lambda_i\Bigg) \\
&\quad + T\bigg(\frac{1}{T_0^{\delta''}}+\delta'\bigg)c_0''\|\boldsymbol{Q}(t_0)\|_1 + Tc_0' \\
&\overset{(b)}{<} \frac{16}{9}(4\epsilon+\delta')TN_T\max_i\{\lambda_i\} \\
&\quad \times\Bigg(\mathbb{E}\Big[\sum_m \frac{\sum_{n=1}^{N^m}Q_m^1(t_0)+N_T(T-t_0)C_A}{\min_{i,m}\{\widetilde{\mu}_{i,m}(t)\}\cdot\min_i\{\widetilde{\mu}_{i,m}(t)\}}\Big|\boldsymbol{Z}(t_0),E\Big]\Bigg) \\
&\quad + T\bigg(\frac{1}{T_0^{\delta''}}+\delta'\bigg)c_0''\|\boldsymbol{Q}(t_0)\|_1 + Tc_0' \\
&\overset{(c)}{<} \bigg(\epsilon+\delta'+\frac{1}{T_0^{\delta''}}\bigg)Tc_1\|\boldsymbol{Q}(t_0)\|_1 + c_0 = \theta_0 T\|\boldsymbol{Q}(t_0)\|_1 + c_0,
\end{aligned}
$$

where (a) follows by upper bounding $1-\epsilon$, $\frac{1}{(1-\epsilon)^2(1+\epsilon)^2}$, and $\frac{1}{t^{\delta''}}$ by 1, $\frac{16}{9}$, and $\frac{1}{T_0^{\delta''}}$, respectively, and (b) is true by the fact that the number of arriving tasks is bounded by C_A, the number of task types is N_T, and the maximum

arrival rate of task types, $\max_i\{\lambda_i\}$, is bounded by the number of servers. Inequality (c) is true by doing simple calculations and using the fact that $\min_{i,m}\{\widetilde{\mu}_{i,m}(t)\}$ is lower bounded by a constant for any $t \geq t_0$ as discussed in (f) of (D.24).

Remark 12. *θ_0 can be made arbitrary small by choosing ϵ and δ' small and T_0 large enough.*

□

D.11 Proof of Lemma 10

Lemma 10. Under the exploration-exploitation scheduling policy of the Blind GB-PANDAS algorithm, for any arrival rate vector inside the capacity region, $\boldsymbol{\lambda} \in \Lambda$, and the corresponding ideal workload vector $\boldsymbol{w}$ in (5.24), there exists $T_1 > 0$ such that for any $T > T_1$, we have:

$$\begin{aligned} &\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\ &\leq -\theta_1 T\|\boldsymbol{Q}(t_0)\|_1 + c_1,\ \forall t_0 \geq 0, \end{aligned} \tag{D.26}$$

where the constants $\theta_1, c_1 > 0$ are independent of $\boldsymbol{Z}(t_0)$.

Proof. The proof is similar to the proof of Lemma 4 and is presented for the sake of completeness. By the assumption on boundedness of arrival and service processes, there exists a constant C_A such that for any t_0, t, and T with $t_0 \leq t \leq t_0 + T$, we have the following for all $m \in \mathcal{M}$:

$$W_m(t_0) - \frac{T}{\min_n\{\alpha_m^n\}} \leq W_m(t) \leq W_m(t_0) + \frac{TC_A}{\min_n\{\alpha_m^n\}}. \tag{D.27}$$

On the other hand, by Lemma 7, the ideal workload on a server defined in (5.24) can be bounded as follows:

$$w_m \leq \frac{1}{1+\delta},\ \forall m \in \mathcal{M}. \tag{D.28}$$

Hence,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&= \mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\sum_{m=1}^{M} W_m(t) w_m\right)\Big|\boldsymbol{Z}(t_0)\right] \\
&\overset{(a)}{\leq} T\sum_{m=1}^{M}\Big(W_m(t_0)w_m + \frac{MT^2C_A}{\min_n\{\alpha_m^n\}}\Big) \\
&\overset{(b)}{\leq} \frac{T}{1+\delta}\sum_m W_m(t_0) + \frac{MT^2C_A}{\min_{m,n}\{\alpha_m^n\}},
\end{aligned} \tag{D.29}$$

where (a) is true by bringing the inner summation on m out of the expectation and using the boundedness property of the workload in Equation (D.27), and (b) is true by Equation (D.28).

Before investigating the second term on the left-hand side of Equation (D.26), $\mathbb{E}\Big[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\Big]$, we propose the following lemma which will be used in lower bounding this second term.

Lemma 14. *For any server $m \in \mathcal{M}$ and any t_0, we have the following:*

$$\lim_{T\to\infty}\frac{\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1}+\frac{S_m^2(t)}{\alpha_m^2}+\cdots+\frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\Big|\boldsymbol{Z}(t_0)\right]}{T} = 1.$$

The proof of Lemma 14 is provided in Appendix D.14. We then have the following:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&= \mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\sum_{m=1}^{M}\left(W_m(t)\left(\frac{S_m^1(t)}{\alpha_m^1}+\frac{S_m^2(t)}{\alpha_m^2}+\cdots+\frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\right)\Big|\boldsymbol{Z}(t_0)\right] \\
&\overset{(a)}{\geq} \sum_{m=1}^{M}\left(W_m(t_0)\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1}+\frac{S_m^2(t)}{\alpha_m^2}+\cdots+\frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\Big|\boldsymbol{Z}(t_0)\right]\right) \\
&\quad -\sum_{m=1}^{M}\left(\frac{T}{\min_n\{\alpha_m^n\}}\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1}+\frac{S_m^2(t)}{\alpha_m^2}+\cdots+\frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\Big|\boldsymbol{Z}(t_0)\right]\right),
\end{aligned} \tag{D.30}$$

where (a) follows by bringing the inner summation on m out of the expectation and using the boundedness property of the workload in Equation (D.27).

Using Lemma 14, for any $0 < \epsilon_0 < \frac{\delta}{1+\delta}$, there exists T_1 such that for any $T \geq T_1$, we have the following for any server $m \in \mathcal{M}$:

$$1 - \epsilon_0 \leq \frac{\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1} + \frac{S_m^2(t)}{\alpha_m^2} + \cdots + \frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\Big|\boldsymbol{Z}(t_0)\right]}{T} \leq 1 + \epsilon_0.$$

Then continuing on Equation (D.30), we have the following:

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right]\\ &\geq T(1-\epsilon_0)\sum_{m=1}^{M} W_m(t_0) - \frac{MT^2(1+\epsilon_0)}{\min_{m,n}\{\alpha_m^n\}}.\end{aligned} \tag{D.31}$$

Then Lemma 10 is concluded as follows by using equations (D.29) and (D.31) and picking $c_1 = \frac{MT^2}{\min_{m,n}\{\alpha_m^n\}}(C_A+1+\epsilon_0)$ and $\theta_1 = \frac{1}{\max_{m,n}\{\alpha_m^n\}}\left(\frac{\delta}{1+\delta} - \epsilon_0\right)$, where by our choice of ϵ_0 we have $\theta_1 > 0$:

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right]\\ &\leq -T\left(\frac{\delta}{1+\delta} - \epsilon_0\right)\sum_{m=1}^{M} W_m(t_0) + \frac{MT^2}{\min_{m,n}\{\alpha_m^n\}}(C_A + 1 + \epsilon_0)\\ &\overset{(a)}{\leq} -\frac{T}{\max_{m,n}\{\alpha_m^n\}}\left(\frac{\delta}{1+\delta} - \epsilon_0\right)\sum_{m=1}^{M}\Big(Q_m^1(t_0) + Q_m^2(t_0)\\ &\quad + \cdots + Q_m^{N^m}(t_0)\Big) + c_1\\ &\leq -\theta_1 T\|\boldsymbol{Q}(t_0)\|_1 + c_1,\ \forall T \geq T_0,\end{aligned}$$

where (a) is true as $W_m(t_0) \geq \frac{Q_m^1(t_0)+Q_m^2(t_0)+\cdots+Q_m^{N^m}(t_0)}{\max_{m,n}\{\alpha_m^n\}}$. □

D.12 Proof of Lemma 11

Lemma 11. Under the exploration-exploitation load balancing of the Blind GB-PANDAS algorithm, for any arrival rate vector inside the capacity region, $\boldsymbol{\lambda} \in \Lambda$, and for any $\theta_2 > 0$, there exists $T_2 > 0$ such that for any $T > T_2$ and

for any $t_0 \geq 0$, we have:

$$\begin{aligned}&\mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0+T)\|_1 - \|\boldsymbol{\Psi}(t_0)\|_1\Big|\boldsymbol{Z}(t_0)\Big]\\ \leq& -\theta_2\|\boldsymbol{\Psi}(t_0)\|_1 + MT.\end{aligned}$$

Proof. For any server $m \in \mathcal{M}$, let t_m^* be the first time slot after or at time slot t_0 at which the server is available; that is,

$$t_m^* = \min\{\tau : \tau \geq t_0, \Psi_m(\tau) = 0\}, \tag{D.32}$$

where it is obvious that $\Psi_m(t_m^*) = 0$. Note that for any $t \geq t_0$, we have $\Psi_m(t+1) \leq \Psi_m(t) + 1$, which is true by the definition of $\Psi(t)$ that is the number of time slots that server m has spent on the currently in-service task. From time slot t to $t+1$, if a new task comes in service, then $\Psi_m(t+1) = 0$ which results in $\Psi_m(t+1) \leq \Psi_m(t)+1$; otherwise, if server m continues giving service to the same task, then $\Psi_m(t+1) = \Psi_m(t) + 1$. Thus, if $t_m^* \leq t_0 + T$, it is easy to find out that $\Psi_m(t_0+T) \leq t_0 + T - t_m^* \leq T$. In the following, we use t_m^* to find a bound on $\mathbb{E}[\Psi_m(t_0+T) - \Psi_m(t_0)|\boldsymbol{Z}(t_0)]$:

$$\begin{aligned}&\mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0+T)\|_1 - \|\boldsymbol{\Psi}(t_0)\|_1\Big|\boldsymbol{Z}(t_0)\Big]\\ =& \sum_{m=1}^{M}\mathbb{E}\left[\Big(\Psi_m(t_0+T) - \Psi_m(t_0)\Big)\Big|\boldsymbol{Z}(t_0)\right]\\ =& \sum_{m=1}^{M}\Bigg\{\mathbb{E}\left[\Big(\Psi_m(t_0+T) - \Psi_m(t_0)\Big)\Big|\boldsymbol{Z}(t_0), t_m^* \leq t_0+T\right]\\ &\times P\left(t_m^* \leq t_0+T\Big|\boldsymbol{Z}(t_0)\right)\\ &+\mathbb{E}\left[\Big(\Psi_m(t_0+T) - \Psi_m(t_0)\Big)\Big|\boldsymbol{Z}(t_0), t_m^* > t_0+T\right]\\ &\times P\left(t_m^* > t_0+T\Big|\boldsymbol{Z}(t_0)\right)\Bigg\}\\ \overset{(a)}{\leq}& \sum_{m=1}^{M}\Bigg\{\Big(T - \Psi_m(t_0)\Big)\times P\left(t_m^* > t_0+T\Big|\boldsymbol{Z}(t_0)\right)\\ &+T\times P\left(t_m^* > t_0+T\Big|\boldsymbol{Z}(t_0)\right)\Bigg\}\\ =& -\sum_{m=1}^{M}\left(\Psi_m(t_0)\cdot P\left(t_m^* > t_0+T\Big|\boldsymbol{Z}(t_0)\right)\right) + MT,\end{aligned} \tag{D.33}$$

where (a) is true as given that $t_m^* \leq t_0 + T$ we found that $\Psi_m(t_0 + T) \leq T$, so $\Psi_m(t_0 + T) - \Psi_m(t_0) \leq T - \Psi_m(t_0)$, and given that $t_m^* > t_0 + T$, it is concluded that server m is giving service to the same task over the whole interval $[t_0, t_0 + T]$, which results in $\Psi_m(t_0 + T) - \Psi_m(t_0) = T$.

Since the CDF of service time of an n-local task on server m has finite mean, we have the following:

$$\lim_{T\to\infty} P\left(t_m^* \leq t_0 + T \Big| \boldsymbol{Z}(t_0)\right) = 1, \ \forall m \in \mathcal{M}.$$

Therefore, for any $\theta_2 \in (0, 1)$ there exists T_2 such that for any $T \geq T_2$, we have $P\left(t_m^* \leq t_0 + T \Big| \boldsymbol{Z}(t_0)\right) \geq \theta_2$, for any $m \in \mathcal{M}$, so Equation (D.33) follows as below which completes the proof:

$$\begin{aligned} &\mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0 + T)\|_1 - \|\boldsymbol{\Psi}(t_0)\|_1 \Big| \boldsymbol{Z}(t_0)\Big] \\ \leq &- \theta_2 \sum_{m=1}^{M} \Psi_m(t_0) + MT = -\theta_2 \|\boldsymbol{\Psi}(t_0)\|_1 + MT. \end{aligned} \tag{D.34}$$

□

D.13 Proof of Lemma 12

Lemma 12. For any $t_0 \leq T_0 < T$, specifically T_0 from Lemma 9 that is dictated by choosing $\theta_0 < \theta_1$, we have the following for the drift of the Lyapunov function in (5.27), where T_0 is used in the first summation after the inequality:

$$\begin{aligned} &\mathbb{E}\Big[V(\boldsymbol{Z}(t_0 + T)) - V(\boldsymbol{Z}(t_0)) \Big| \boldsymbol{Z}(t_0)\Big] \\ \leq\ &2\mathbb{E}\left[\sum_{t=T_0}^{t_0+T-1} \Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big) \Big| \boldsymbol{Z}(t_0)\right] \\ &+ 2\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1} \Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big) \Big| \boldsymbol{Z}(t_0)\right] \\ &+ \mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0 + T)\|_1 - \|\boldsymbol{\Psi}(t)\|_1 \Big| \boldsymbol{Z}(t_0)\Big] + c_2\|\boldsymbol{Q}(t_0)\|_1 + c_3. \end{aligned} \tag{D.35}$$

Proof.

$$
\begin{aligned}
&\mathbb{E}\Big[V(\boldsymbol{Z}(t_0+T)) - V(\boldsymbol{Z}(t_0))\Big|\boldsymbol{Z}(t_0)\Big] \\
&= \mathbb{E}\Big[\|\boldsymbol{W}(t_0+T)\|^2 - \|\boldsymbol{W}(t_0)\|^2\Big|\boldsymbol{Z}(t_0)\Big] \\
&\quad + \mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0+T)\|_1 - \|\boldsymbol{\Psi}(t_0)\|_1\Big|\boldsymbol{Z}(t_0)\Big] \\
&\overset{(a)}{=} \mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\|\boldsymbol{W}(t+1)\|^2 - \|\boldsymbol{W}(t)\|^2\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\quad + \mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0+T)\|_1 - \|\boldsymbol{\Psi}(t)\|_1\Big|\boldsymbol{Z}(t_0)\Big] \\
&\overset{(b)}{=} \mathbb{E}\Bigg[\sum_{t=t_0}^{t_0+T-1}\Big(\|\boldsymbol{A}(t) - \boldsymbol{S}(t) + \widetilde{\boldsymbol{U}}(t)\|^2 \\
&\quad + 2\langle\boldsymbol{W}(t), \boldsymbol{A}(t) - \boldsymbol{S}(t)\rangle + 2\langle\boldsymbol{W}(t), \widetilde{\boldsymbol{U}}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\Bigg] \\
&\quad + \mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0+T)\|_1 - \|\boldsymbol{\Psi}(t)\|_1\Big|\boldsymbol{Z}(t_0)\Big] \\
&\overset{(c)}{\leq} 2\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle\boldsymbol{W}(t), \boldsymbol{A}(t) - \boldsymbol{S}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\quad + \mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0+T)\|_1 - \|\boldsymbol{\Psi}(t)\|_1\Big|\boldsymbol{Z}(t_0)\Big] + c_3' \\
&\overset{(d)}{=} 2\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle\boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle\boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\quad + 2\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle\boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle\boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\quad + \mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0+T)\|_1 - \|\boldsymbol{\Psi}(t)\|_1\Big|\boldsymbol{Z}(t_0)\Big] + c_3',
\end{aligned}
\tag{D.36}
$$

where (a) is true by the telescoping series, (b) follows by using (5.22) to substitute $\boldsymbol{W}(t+1)$, (c) follows by Lemma 8 and the fact that the task arrival is assumed to be bounded and the service and unused service are also bounded as the number of servers are finite, so the pseudo arrival, service, and unused service are also bounded, and therefore there exists a constant c_1 such that $\|\boldsymbol{A}(t) - \boldsymbol{S}(t) + \widetilde{\boldsymbol{U}}(t)\|^2 \leq \frac{c_3'}{T}$, and (d) follows by adding and

subtracting the intermediary term $\langle \boldsymbol{W}(t), \boldsymbol{w} \rangle$. On the other hand,

$$
\begin{aligned}
&2\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\leq 2\mathbb{E}\left[\sum_{t=t_0}^{T_0-1}\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle\Big|\boldsymbol{Z}(t_0)\right] \\
&\quad + 2\mathbb{E}\left[\sum_{t=T_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\overset{(a)}{\leq} 2\mathbb{E}\Bigg[\frac{(T_0-t_0)\cdot C_A}{\left(\min_{m,n}\{\alpha_m^n\}\right)^2}\sum_{m\in\mathcal{M}}\big(Q_m^1(t_0)+\cdots+Q_m^{N^m}(t_0) \\
&\quad + N^m\cdot C_A\cdot(T_0-t_0)\big)\Big|\boldsymbol{Z}(t_0)\Bigg] \\
&\quad + 2\mathbb{E}\left[\sum_{t=T_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\leq 2\mathbb{E}\left[\sum_{t=T_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] + c_2\|\boldsymbol{Q}(t_0)\|_1 + c_3'',
\end{aligned}
\tag{D.37}
$$

where (a) is true by the fact that at most C_A tasks arrive at system in each time slot, and by using the definition of pseudo task arrival in (5.21). Putting (D.36) and (D.37) together, Lemma 12 is proved as follows:

$$
\begin{aligned}
&\mathbb{E}\Big[V(\boldsymbol{Z}(t_0+T)) - V(\boldsymbol{Z}(t_0))\Big|\boldsymbol{Z}(t_0)\Big] \\
&\leq 2\mathbb{E}\left[\sum_{t=T_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{A}(t)\rangle - \langle \boldsymbol{W}(t), \boldsymbol{w}\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\quad + 2\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\Big(\langle \boldsymbol{W}(t), \boldsymbol{w}\rangle - \langle \boldsymbol{W}(t), \boldsymbol{S}(t)\rangle\Big)\Big|\boldsymbol{Z}(t_0)\right] \\
&\quad + \mathbb{E}\Big[\|\boldsymbol{\Psi}(t_0+T)\|_1 - \|\boldsymbol{\Psi}(t)\|_1\Big|\boldsymbol{Z}(t_0)\Big] + c_2\|\boldsymbol{Q}(t_0)\|_1 + c_3,
\end{aligned}
$$

where $c_3 = c_3' + c_3''$. □

D.14 Proof of Lemma 14

Lemma 14. For any server $m \in \mathcal{M}$ and any t_0, we have the following:

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1}+\frac{S_m^2(t)}{\alpha_m^2}+\cdots+\frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\middle|\boldsymbol{Z}(t_0)\right]}{T} = 1.$$

Proof. The proof is similar to the proof of Lemma 13 and is presented for the sake of completeness. Let t_m^* be the first time slot after or at time slot t_0 at which server m becomes idle, and so is available to serve another task (t_m^* is also defined in (D.32)); that is,

$$t_m^* = \min\{\tau : \tau \geq t_0, \Psi_m(\tau) = 0\}, \tag{D.38}$$

where, as a reminder, $\Psi_m(\tau)$ is the number of time slots that the m-th server has spent on the task that is receiving service from this server at time slot τ.

Denote the CDF of service time of an n-local task on server m by F_m^n that has finite mean $\alpha_m^n < \infty$; therefore, $t_m^* < \infty$. We then have the following by considering the bounded service:

$$\begin{aligned}
&\left(\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1}+\cdots+\frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\middle|\boldsymbol{Z}(t_0)\right] - \frac{t_m^*-t_0}{\alpha_m^{N^m}} + \frac{1}{\alpha_m^1}\right)\Bigg/T \\
&\leq \frac{\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1}+\frac{S_m^2(t)}{\alpha_m^2}+\cdots+\frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\middle|\boldsymbol{Z}(t_0)\right]}{T} \\
&\leq \left(\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1}+\cdots+\frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\middle|\boldsymbol{Z}(t_0)\right] + \frac{1}{\alpha_m^{N^m}}\right)\Bigg/T,
\end{aligned} \tag{D.39}$$

where by boundedness of t_m^*, α_m^1, and $\alpha_m^{N^m}$, it is obvious that

$$\lim_{T\to\infty}\frac{-\frac{t_m^*-t_0}{\alpha_m^{N^m}}+\frac{1}{\alpha_m^1}}{T} = 0 \quad \text{and} \quad \lim_{T\to\infty}\frac{\frac{1}{\alpha_m^{N^m}}}{T} = 0.$$

Hence, by taking the limit of the terms in Equation (D.39) as T goes to

infinity, we have the following:

$$\begin{aligned}&\lim_{T\to\infty}\frac{\mathbb{E}\left[\sum_{t=t_0}^{t_0+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1}+\frac{S_m^2(t)}{\alpha_m^2}+\cdots+\frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\middle|\boldsymbol{Z}(t_0)\right]}{T}\\ =&\lim_{T\to\infty}\frac{\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1}+\frac{S_m^2(t)}{\alpha_m^2}+\cdots+\frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\middle|\boldsymbol{Z}(t_0)\right]}{T}.\end{aligned}\tag{D.40}$$

Considering the service process as a renewal process, given the scheduling decisions at the end of the renewal intervals in $[t_m^*, t_m^*+T-1]$, all holding times for server m to give service to tasks in its sub-queues are independent. We elaborate on this in the following. We define renewal processes, $N_m^n(t),\ n \in \{1, 2, \cdots, N^m\}$, as follows, where t is an integer valued number:
Let $H_m^n(l)$ be the holding time (service time) of the l-th task that is n-local to server m after time slot t_m^* receiving service from server m, and call $\{H_m^n(l), l \geq 1\}$ the holding process of n-local task type, $n \in \{1, 2, \cdots, N^m\}$. Then define $J_m^n(l) = \sum_{i=1}^{l} H_m^n(l)$ for $l \geq 1$, and let $J_m^n(0) = 0$. In the renewal process, $J_m^n(l)$ is the l-th jumping time, or the time at which the l-th occurrence happens, and it has the following relation with the renewal process, $N_m^n(t)$:

$$N_m^n(t) = \sum_{l=1}^{\infty} \mathbb{I}_{\{J_m^n(l)\leq t\}} = \sup\{l : J_m^n(l) \leq t\}.$$

Another way to define $N_m^n(t)$ is as shown in the following algorithm, where by convention, $N_m^n(0) = 0$.

Set $\tau = t_m^*$, $cntr = 0$, $N_m^n(t) = 0$
while $cntr < t$ **do**
  if $\eta_m(\tau) = n$ **then**
    $cntr++$
    $N_m^n(t) += S_m^n(\tau)$
  end if
  $\tau++$
end while

Another renewal process, $N_m(t)$, is defined as

$$N_m(t) = \sum_{u=t_m^*}^{t_m^*+t-1} \Big(\mathbb{I}_{\{S_m^1(u)=1\}} + \mathbb{I}_{\{S_m^2(u)=1\}} + \cdots + \mathbb{I}_{\{S_m^{N^m}(u)=1\}} \Big).$$

Similarly, let $H_m(l)$ be the holding time (service time) of the l-th task after time slot t_m^* receiving service from server m, and call $\{H_m(l), l \geq 1\}$ the holding process. Then define $J_m(l) = \sum_{i=1}^{l} H_m(l)$ for $l \geq 1$, and let $J_m(0) = 0$. In the renewal process, $J_m(l)$ is the l-th jumping time, or the time at which the l-th occurrence happens, and it has the following relation with the renewal process, $N_m(t)$:

$$N_m(t) = \sum_{l=1}^{\infty} \mathbb{I}_{\{J_m(l) \leq t\}} = \sup\{l : J_m(l) \leq t\}.$$

Note that the central scheduler makes scheduling decisions for server m at time slots $\{t_m^* + J_m(l), l \geq 1\}$. We denote these scheduling decisions by $D_m(t_m^*) = \Big(\eta_m(t_m^* + J_m(l)) : l \geq 1\Big)$. Consider the time interval $[t_m^*, t_m^* + T - 1]$ when T goes to infinity. Define ρ_m^n as the fraction of time that server m is busy giving service to tasks that are n-local to this server, in the mentioned interval. Obviously, $\sum_{n=1}^{N^m} \rho_m^n = 1$. Then Equation (D.40) is followed by

$$\begin{aligned}
&\lim_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1} \left(\frac{S_m^1(t)}{\alpha_m^1} + \frac{S_m^2(t)}{\alpha_m^2} + \cdots + \frac{S_m^{N^m}(t)}{\alpha_m^{N^m}} \right) \middle| \boldsymbol{Z}(t_0) \right]}{T} \\
&= \lim_{T\to\infty} \Bigg\{ \mathbb{E}\Bigg[\mathbb{E}\Bigg[\sum_{t=t_m^*}^{t_m^*+T-1} \Bigg(\frac{S_m^1(t)}{\alpha_m^1} + \frac{S_m^2(t)}{\alpha_m^2} \\
&\quad + \cdots + \frac{S_m^{N^m}(t)}{\alpha_m^{N^m}} \Bigg) \Bigg| D_m(t_m^*), \boldsymbol{Z}(t_0) \Bigg] \Bigg| \boldsymbol{Z}(t_0) \Bigg] \Bigg\} \Bigg/ T \\
&= \sum_{n=1}^{N^m} \lim_{T\to\infty} \left(\mathbb{E}\left[\frac{1}{\alpha_m^n} \mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1} \Big(S_m^n(t) \Big) \middle| D_m(t_m^*), \boldsymbol{Z}(t_0) \right] \middle| \boldsymbol{Z}(t_0) \right] \right) \Bigg/ T \\
&= \sum_{n=1}^{N^m} \mathbb{E}\left[\frac{1}{\alpha_m^n} \lim_{T\to\infty} \frac{\mathbb{E}\left[N_m^n\big(\rho_m^n T\big) \middle| D_m(t_m^*), \boldsymbol{Z}(t_0) \right]}{T} \middle| \boldsymbol{Z}(t_0) \right].
\end{aligned} \tag{D.41}$$

Note that given $\{D_m(t_m^*), \boldsymbol{Z}(t_0)\}$, the holding times $\{H_m^n(l), l \geq 1\}$ are independent and identically distributed with CDF F_m^n. If $\rho_m^n = 0$, then we

do not have to worry about those tasks that are n-local to server m since they receive service from this server for only a finite number of times in time interval $[t_m^*, t_m^* + T - 1]$ as $T \to \infty$, so

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[N_m^n\left(\rho_m^n T\right)\middle| D_m(t_m^*), \boldsymbol{Z}(t_0)\right]}{T} = 0.$$

But if $\rho_m^n > 0$, we can use the strong law of large numbers for renewal process N_m^n to conclude the following:

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[N_m^n\left(\rho_m^n T\right)\middle| D_m(t_m^*), \boldsymbol{Z}(t_0)\right]}{T} = \rho_m^n \cdot \frac{1}{\mathbb{E}[H_m^n(1)]}, \tag{D.42}$$

where the holding time (service time) $H_m^n(1)$ has CDF F_m^n with expectation $\frac{1}{\alpha_m^n}$. Combining equations (D.41) and (D.42), Lemma 14 is concluded as follows:

$$\begin{aligned} &\lim_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=t_m^*}^{t_m^*+T-1}\left(\frac{S_m^1(t)}{\alpha_m^1} + \frac{S_m^2(t)}{\alpha_m^2} + \cdots + \frac{S_m^{N^m}(t)}{\alpha_m^{N^m}}\right)\middle| \boldsymbol{Z}(t_0)\right]}{T} \\ &= \sum_{n=1}^{N^m} \mathbb{E}\left[\frac{1}{\alpha_m^n} \cdot \rho_m^n \cdot \alpha_m^n \middle| \boldsymbol{Z}(t_0)\right] = \sum_{n=1}^{N^m} \rho_m^n = 1. \end{aligned} \tag{D.43}$$

□

BIBLIOGRAPHY

[1] L. X. Bui, R. Johari, and S. Mannor, "Committing bandits," in *Advances in Neural Information Processing Systems*, 2011, pp. 1557–1565.

[2] A. Garivier, T. Lattimore, and E. Kaufmann, "On explore-then-commit strategies," in *Advances in Neural Information Processing Systems*, 2016, pp. 784–792.

[3] D. Liau, E. Price, Z. Song, and G. Yang, "Stochastic multi-armed bandits in constant space," *arXiv preprint arXiv:1712.09007*, 2017.

[4] V. Perchet, P. Rigollet, S. Chassang, E. Snowberg et al., "Batched bandit problems," *The Annals of Statistics*, vol. 44, no. 2, pp. 660–681, 2016.

[5] L. Tran-Thanh, A. Chapman, A. Rogers, and N. R. Jennings, "Knapsack based optimal policies for budget–limited multi–armed bandits," in *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[6] W. Ding, T. Qin, X.-D. Zhang, and T.-Y. Liu, "Multi-armed bandit with budget constraint and variable costs," in *Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.

[7] L. Tran-Thanh, A. Chapman, E. M. de Cote, A. Rogers, and N. R. Jennings, "Epsilon–first policies for budget–limited multi-armed bandits," in *Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[8] Y. Xia, W. Ding, X.-D. Zhang, N. Yu, and T. Qin, "Budgeted bandit problems with continuous random costs," in *Asian Conference on Machine Learning*, 2016, pp. 317–332.

[9] Y. Xia, H. Li, T. Qin, N. Yu, and T.-Y. Liu, "Thompson sampling for budgeted multi-armed bandits," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[10] A. Badanidiyuru, R. Kleinberg, and A. Slivkins, "Bandits with knapsacks," in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 2013, pp. 207–216.

[11] S. Agrawal and N. Devanur, "Linear contextual bandits with knapsacks," in *Advances in Neural Information Processing Systems*, 2016, pp. 3450–3458.

[12] A. Badanidiyuru, J. Langford, and A. Slivkins, "Resourceful contextual bandits," in *Conference on Learning Theory*, 2014, pp. 1109–1134.

[13] S. Bubeck, R. Munos, and G. Stoltz, "Pure exploration in finitely-armed and continuous-armed bandits," *Theoretical Computer Science*, vol. 412, no. 19, pp. 1832–1852, 2011.

[14] J.-Y. Audibert and S. Bubeck, "Best arm identification in multi-armed bandits," in *COLT-23th Conference on learning theory-2010*, 2010, pp. 13–p.

[15] V. Gabillon, M. Ghavamzadeh, A. Lazaric, and S. Bubeck, "Multi-bandit best arm identification," in *Advances in Neural Information Processing Systems*, 2011, pp. 2222–2230.

[16] H. M. Markowitz, "Portfolio selection/Harry Markowitz," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[17] A. Sani, A. Lazaric, and R. Munos, "Risk-aversion in multi-armed bandits," in *Advances in Neural Information Processing Systems*, 2012, pp. 3275–3283.

[18] S. Vakili and Q. Zhao, "Risk-averse multi-armed bandit problems under mean-variance measure," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 6, pp. 1093–1111, 2016.

[19] S. Vakili and Q. Zhao, "Mean-variance and value at risk in multi-armed bandit problems," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2015, pp. 1330–1335.

[20] J. Y. Yu and E. Nikolova, "Sample complexity of risk-averse bandit-arm selection," in *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.

[21] S. Vakili and Q. Zhao, "Risk-averse online learning under mean-variance measures," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 1911–1915.

[22] N. Galichet, M. Sebag, and O. Teytaud, "Exploration vs exploitation vs safety: Risk-aware multi-armed bandits," in *Asian Conference on Machine Learning*, 2013, pp. 245–260.

[23] J. Xu, W. B. Haskell, and Z. Ye, "Index-based policy for risk-averse multi-armed bandit," *arXiv preprint arXiv:1809.05385*, 2018.

[24] N. Galichet, "Contributions to multi-armed bandits: Risk-awareness and sub-sampling for linear contextual bandits," Ph.D. dissertation, Université Paris Sud-Paris XI, 2015.

[25] A. Cassel, S. Mannor, and A. Zeevi, "A general approach to multi-armed bandits under risk criteria," *arXiv preprint arXiv:1806.01380*, 2018.

[26] R. K. Kolla, K. Jagannathan et al., "Risk-aware multi-armed bandits using conditional value-at-risk," *arXiv preprint arXiv:1901.00997*, 2019.

[27] J. von Neumann, "Zur theorie der gesellschaftsspiele," *Mathematische Annalen*, vol. 100, no. 1, pp. 295–320, Dec 1928. [Online]. Available: https://doi.org/10.1007/BF01448847

[28] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton University Press, 1947.

[29] J. F. Nash, "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences*, vol. 36, no. 1, pp. 48–49, 1950. [Online]. Available: https://www.pnas.org/content/36/1/48

[30] J. C. Harsanyi, "Games with incomplete information played by "Bayesian" players, I-III part I. the basic model," *Management Science*, vol. 14, 1967. [Online]. Available: https://doi.org/10.1287/mnsc.14.3.159

[31] T. Wiseman, "A partial folk theorem for games with unknown payoff distributions," *Econometrica*, vol. 73, no. 2, pp. 629–645, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2005.00589.x

[32] P. Mertikopoulos and Z. Zhou, "Learning in games with continuous action sets and unknown payoff functions," *Mathematical Programming*, vol. 173, no. 1, pp. 465–507, Jan 2019. [Online]. Available: https://doi.org/10.1007/s10107-018-1254-8

[33] T. Sugaya and Y. Yamamoto, "Common learning and cooperation in repeated games," 2019. [Online]. Available: https://dx.doi.org/10.2139/ssrn.3385516

[34] A. Yekkehkhany, T. Murray, and R. Nagi, "Risk-averse equilibrium for games," *arXiv preprint arXiv:2002.08414*, 2020.

[35] A. Yekkehkhany and R. Nagi, "Risk-averse equilibrium for autonomous vehicles in stochastic congestion games."

[36] H. Angelidakis, D. Fotakis, and T. Lianeas, "Stochastic congestion games with risk-averse players," in *Algorithmic Game Theory*, B. Vöcking, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 86–97.

[37] M. G. Bell and C. Cassir, "Risk-averse user equilibrium traffic assignment: An application of game theory," *Transportation Research Part B: Methodological*, vol. 36, no. 8, pp. 671 – 681, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0191261501000224

[38] T. Yamazaki, "The uniqueness of pure-strategy Nash equilibrium in rent-seeking games with risk-averse players," *Public Choice*, vol. 139, no. 3/4, pp. 335–342, 2009. [Online]. Available: http://www.jstor.org/stable/40270893

[39] J. E. Harrington, "A non-cooperative bargaining game with risk averse players and an uncertain finite horizon," *Economics Letters*, vol. 20, no. 1, pp. 9 – 13, 1986. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0165176586900704

[40] J. K. Goeree, C. A. Holt, and T. R. Palfrey, "Risk averse behavior in generalized matching pennies games," *Games and Economic Behavior*, vol. 45, no. 1, pp. 97 – 113, 2003, first World Congress of the Game Theory Society. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0899825603000526

[41] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, pp. 263–292, 1979.

[42] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, Oct 1992. [Online]. Available: https://doi.org/10.1007/BF00122574

[43] J. S. Levy, "An introduction to prospect theory," *Political Psychology*, vol. 13, no. 2, pp. 171–186, 1992. [Online]. Available: http://www.jstor.org/stable/3791677

[44] L. Baele, J. Driessen, S. Ebert, J. M. Londono, and O. G. Spalt, "Cumulative prospect theory, option returns, and the variance premium," *The Review of Financial Studies*, vol. 32, no. 9, pp. 3667–3723, 12 2018. [Online]. Available: https://doi.org/10.1093/rfs/hhy127

[45] N. Barberis, A. Mukherjee, and B. Wang, "Prospect theory and stock returns: An empirical test," *The Review of Financial Studies*, vol. 29, no. 11, pp. 3068–3107, 07 2016. [Online]. Available: https://doi.org/10.1093/rfs/hhw049

[46] N. Barberis, M. Huang, and T. Santos, "Prospect theory and asset prices," *The Quarterly Journal of Economics*, vol. 116, no. 1, pp. 1–53, 02 2001. [Online]. Available: https://doi.org/10.1162/003355301556310

[47] J. A. List, "Neoclassical theory versus prospect theory: Evidence from the marketplace," *Econometrica*, vol. 72, no. 2, pp. 615–625, 2004. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2004.00502.x

[48] R. Bellman, "On a routing problem," *Quarterly of Applied Mathematics*, vol. 16, no. 1, pp. 87–90, 1958.

[49] E. Dijkstra, "A note on two problems in connection with graphs," *Numerical Mathematics*, vol. 1, pp. 269–271, 1959.

[50] S. E. Dreyfus, "An appraisal of some shortest-path algorithms," *Operations Research*, vol. 17, no. 3, pp. 395–412, 1969.

[51] A. Schrijver, "On the history of the shortest path problem," *Documenta Mathematica*, vol. 17, p. 155, 2012.

[52] L. Fu, D. Sun, and L. R. Rilett, "Heuristic shortest path algorithms for transportation applications: State of the art," *Computers & Operations Research*, vol. 33, no. 11, pp. 3324–3343, 2006.

[53] R. B. Dial, "Algorithm 360: Shortest-path forest with topological ordering [h]," *Communications of the ACM*, vol. 12, no. 11, pp. 632–633, 1969.

[54] R. E. Tarjan, *Data Structures and Network Algorithms*. SIAM, 1983.

[55] E. Lawler, *Combinatorial Optimization: Networks and Matroids*. New York: Rinehart & Winston, 1976.

[56] A. Pierce, "Bibliography on algorithms for shortest path, shortest spanning tree, and related circuit routing problems (1956–1974)," *Networks*, vol. 5, no. 2, pp. 129–149, 1975.

[57] A. Orda and R. Rom, "Shortest-path and minimum-delay algorithms in networks with time-dependent edge-length," *Journal of the ACM (JACM)*, vol. 37, no. 3, pp. 607–625, 1990.

[58] D. E. Kaufman and R. L. Smith, "Fastest paths in time-dependent networks for intelligent vehicle-highway systems application," *Journal of Intelligent Transportation Systems*, vol. 1, no. 1, pp. 1–11, 1993.

[59] A. K. Ziliaskopoulos and H. S. Mahmassani, "Time-dependent, shortest-path algorithm for real-time intelligent vehicle highway system applications," *Transportation Research Record*, 1993.

[60] I. Chabini, "A new algorithm for shortest paths in discrete dynamic networks," *IFAC Proceedings Volumes*, vol. 30, no. 8, pp. 537–542, 1997.

[61] R. P. Loui, "Optimal paths in graphs with stochastic or multidimensional weights," *Communications of the ACM*, vol. 26, no. 9, pp. 670–676, 1983.

[62] H. Frank, "Shortest paths in probabilistic graphs," *Operations Research*, vol. 17, no. 4, pp. 583–599, 1969.

[63] C. E. Sigal, A. A. B. Pritsker, and J. J. Solberg, "The stochastic shortest route problem," *Operations Research*, vol. 28, no. 5, pp. 1122–1129, 1980.

[64] A. Chen and Z. Ji, "Path finding under uncertainty," *Journal of Advanced Transportation*, vol. 39, no. 1, pp. 19–37, 2005.

[65] Y. Nie and Y. Fan, "Arriving-on-time problem: Discrete algorithm that ensures convergence," *Transportation Research Record*, vol. 1964, no. 1, pp. 193–200, 2006.

[66] Y. M. Nie and X. Wu, "Shortest path problem considering on-time arrival probability," *Transportation Research Part B: Methodological*, vol. 43, no. 6, pp. 597–613, 2009.

[67] W. Zeng, T. Miwa, Y. Wakita, and T. Morikawa, "Application of Lagrangian relaxation approach to α-reliable path finding in stochastic networks with correlated link travel times," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 309–334, 2015.

[68] T. Xing and X. Zhou, "Finding the most reliable path with and without link travel time correlation: A Lagrangian substitution based approach," *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1660–1679, 2011.

[69] R. A. Howard, *Dynamic Probabilistic Systems: Markov Models*. Courier Corporation, 2012, vol. 1.

[70] R. W. Hall, "The fastest path through a network with random time-dependent travel times," *Transportation Science*, vol. 20, no. 3, pp. 182–188, 1986.

[71] L. Fu and L. R. Rilett, "Expected shortest paths in dynamic and stochastic traffic networks," *Transportation Research Part B: Methodological*, vol. 32, no. 7, pp. 499–516, 1998.

[72] S. T. Waller and A. K. Ziliaskopoulos, "On the online shortest path problem with limited arc cost dependencies," *Networks: An International Journal*, vol. 40, no. 4, pp. 216–227, 2002.

[73] E. D. Miller-Hooks and H. S. Mahmassani, "Least expected time paths in stochastic, time-varying transportation networks," *Transportation Science*, vol. 34, no. 2, pp. 198–215, 2000.

[74] B. Mirchandani, H. Soroush, G. Angrealtta, F. Mason, and P. Serafini, "Routes and flows in stochastic networks," *Advanced Schools on Stochastic in Combinatorial Optimization*, pp. 129–177, 1986.

[75] P. B. Mirchandani, "Shortest distance and reliability of probabilistic networks," *Computers & Operations Research*, vol. 3, no. 4, pp. 347–355, 1976.

[76] I. Murthy and S. Sarkar, "A relaxation-based pruning technique for a class of stochastic shortest path problems," *Transportation Science*, vol. 30, no. 3, pp. 220–236, 1996.

[77] Y. Fan, "Optimal routing through stochastic networks," Ph.D. dissertation, University of Southern California, 2003.

[78] L. Xiao and H. K. Lo, "Adaptive vehicle routing for risk-averse travelers," *Procedia-Social and Behavioral Sciences*, vol. 80, pp. 633–657, 2013.

[79] M. G. Bell, "Hyperstar: A multi-path Astar algorithm for risk averse vehicle navigation," *Transportation Research Part B: Methodological*, vol. 43, no. 1, pp. 97–107, 2009.

[80] Y. Chen, M. G. Bell, and K. Bogenberger, "Risk-averse autonomous route guidance by a constrained A* search," *Journal of Intelligent Transportation Systems*, vol. 14, no. 3, pp. 188–196, 2010.

[81] H. K. Lo, X. Luo, and B. W. Siu, "Degradable transport network: Travel time budget of travelers with heterogeneous risk aversion," *Transportation Research Part B: Methodological*, vol. 40, no. 9, pp. 792–806, 2006.

[82] J. G. Wardrop and J. I. Whitehead, "some theoretical aspects of road traffic research," *Proceedings of the Institution of Civil Engineers*, vol. 1, no. 5, pp. 767–768, 1952.

[83] J. v. Neumann, "Zur theorie der gesellschaftsspiele," *Mathematische annalen*, vol. 100, no. 1, pp. 295–320, 1928.

[84] J. Von Neumann and O. Morgenstern, "Theory of Games and Economic Behavior," *2nd ed., Princeton University Press*, 1947.

[85] J. F. Nash et al., "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences*, vol. 36, no. 1, pp. 48–49, 1950.

[86] J. C. Harsanyi, "Games with incomplete information played by "Bayesian" players, I–III part I. the basic model," *Management Science*, vol. 14, no. 3, pp. 159–182, 1967.

[87] J. C. Harsanyi, "Games with incomplete information played by "Bayesian" players part II. Bayesian equilibrium points," *Management Science*, vol. 14, no. 5, pp. 320–334, 1968.

[88] F. Ordóñez and N. E. Stier-Moses, "Wardrop equilibria with risk-averse users," *Transportation Science*, vol. 44, no. 1, pp. 63–86, 2010.

[89] D. Watling, "User equilibrium traffic network assignment with stochastic travel times and late arrival penalty," *European Journal of Operational Research*, vol. 175, no. 3, pp. 1539–1556, 2006.

[90] W. Szeto, L. O'Brien, and M. O'Mahony, "Risk-averse traffic assignment with elastic demands: NCP formulation and solution method for assessing performance reliability," *Networks and Spatial Economics*, vol. 6, no. 3-4, pp. 313–332, 2006.

[91] A. Chen and Z. Zhou, "The α-reliable mean-excess traffic equilibrium model with stochastic travel times," *Transportation Research Part B: Methodological*, vol. 44, no. 4, pp. 493–513, 2010.

[92] M. G. Bell and C. Cassir, "Risk-averse user equilibrium traffic assignment: An application of game theory," *Transportation Research Part B: Methodological*, vol. 36, no. 8, pp. 671–681, 2002.

[93] H. Z. Aashtiani and T. L. Magnanti, "Equilibria on a congested transportation network," *SIAM Journal on Algebraic Discrete Methods*, vol. 2, no. 3, pp. 213–226, 1981.

[94] M. Aghassi and D. Bertsimas, "Robust game theory," *Mathematical Programming*, vol. 107, no. 1-2, pp. 231–273, 2006.

[95] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter, "A survey on networking games in telecommunications," *Computers & Operations Research*, vol. 33, no. 2, pp. 286–311, 2006.

[96] S. Hayashi, N. Yamashita, and M. Fukushima, "Robust Nash equilibria and second-order cone complementarity problems," *Journal of Nonlinear and Convex Analysis*, vol. 6, no. 2, p. 283, 2005.

[97] P. Mirchandani and H. Soroush, "Generalized traffic equilibrium with probabilistic travel times and perceptions," *Transportation Science*, vol. 21, no. 3, pp. 133–152, 1987.

[98] Y. M. Nie, "Multi-class percentile user equilibrium with flow-dependent stochasticity," *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1641–1659, 2011.

[99] R. D. Connors and A. Sumalee, "A network equilibrium model with travellers' perception of stochastic travel times," *Transportation Research Part B: Methodological*, vol. 43, no. 6, pp. 614–624, 2009.

[100] J.-D. Schmöcker, M. G. Bell, F. Kurauchi, and H. Shimamoto, "A game theoretic approach to the determination of hyperpaths in transportation networks," in *Transportation and Traffic Theory 2009: Golden Jubilee*. Springer, 2009, pp. 1–18.

[101] A. Fonzone, J.-D. Schmöcker, J. Ma, and D. Fukuda, "Link-based route choice considering risk aversion, disappointment, and regret," *Transportation Research Record*, vol. 2322, no. 1, pp. 119–128, 2012.

[102] H. Angelidakis, D. Fotakis, and T. Lianeas, "Stochastic congestion games with risk-averse players," in *International Symposium on Algorithmic Game Theory*. Springer, 2013, pp. 86–97.

[103] E. Nikolova and N. E. Stier-Moses, "Stochastic selfish routing," in *International Symposium on Algorithmic Game Theory*. Springer, 2011, pp. 314–325.

[104] E. Nikolova and N. E. Stier-Moses, "The burden of risk aversion in mean-risk selfish routing," in *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, 2015, pp. 489–506.

[105] J. Correa, R. Hoeksma, and M. Schröder, "Network congestion games are robust to variable demand," *Transportation Research Part B: Methodological*, vol. 119, pp. 69–78, 2019.

[106] T. White, *Hadoop: The Definitive Guide*. O'Reilly and Yahoo! Press, 2nd ed., 2010.

[107] M. Isard, V. Prabhakaran, J. Currey, U. Wieder, K. Talwar, and A. Goldberg, "Quincy: Fair scheduling for distributed computing clusters," in *Proceedings of the ACM SIGOPS 22nd Symposium on Operating Systems Principles*. ACM, 2009, pp. 261–276.

[108] M. Zaharia, D. Borthakur, J. Sen Sarma, K. Elmeleegy, S. Shenker, and I. Stoica, "Delay scheduling: A simple technique for achieving locality and fairness in cluster scheduling," in *Proceedings of the 5th European Conference on Computer Systems*. ACM, 2010, pp. 265–278.

[109] J. Jin, J. Luo, A. Song, F. Dong, and R. Xiong, "Bar: An efficient data locality driven task scheduling algorithm for cloud computing," in *Cluster, Cloud and Grid Computing (CCGrid), 2011 11th IEEE/ACM International Symposium on*. IEEE, 2011, pp. 295–304.

[110] C. He, Y. Lu, and D. Swanson, "Matchmaking: A new MapReduce scheduling technique," in *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*. IEEE, 2011, pp. 40–47.

[111] S. Ibrahim, H. Jin, L. Lu, B. He, G. Antoniu, and S. Wu, "Maestro: Replica-aware map scheduling for MapReduce," in *Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on*. IEEE, 2012, pp. 435–442.

[112] J. Polo, C. Castillo, D. Carrera, Y. Becerra, I. Whalley, M. Steinder, J. Torres, and E. Ayguadé, "Resource-aware adaptive scheduling for MapReduce clusters," in *ACM/IFIP/USENIX International Conference on Distributed Systems Platforms and Open Distributed Processing*. Springer, 2011, pp. 187–207.

[113] M. Zaharia, A. Konwinski, A. D. Joseph, R. H. Katz, and I. Stoica, "Improving MapReduce performance in heterogeneous environments." in *Osdi*, vol. 8, no. 4, 2008, p. 7.

[114] M. Hosseini, Y. Jiang, A. Yekkehkhany, R. R. Berlin, and L. Sha, "A mobile geo-communication dataset for physiology-aware dash in rural ambulance transport," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, 2017, pp. 158–163.

[115] J. Jin, Q. An, W. Zhou, J. Tang, and R. Xiong, "Dyndl: Scheduling data-locality-aware tasks with dynamic data transfer cost for multicore-server-based big data clusters," *Applied Sciences*, vol. 8, no. 11, p. 2216, 2018.

[116] W. Tang, X. Liu, W. Rafique, and W. Dou, "A dynamic resource allocation method for load-balance scheduling over big data platforms," in

2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2018, pp. 524–531.

[117] F. Chen, M. Kodialam, and T. Lakshman, "Joint scheduling of processing and shuffle phases in MapReduce systems," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 1143–1151.

[118] J. Tan, X. Meng, and L. Zhang, "Coupling task progress for MapReduce resource-aware scheduling," in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 1618–1626.

[119] M. Lin, L. Zhang, A. Wierman, and J. Tan, "Joint optimization of overlapping phases in MapReduce," *Performance Evaluation*, vol. 70, no. 10, pp. 720–735, 2013.

[120] J. M. Harrison, "Heavy traffic analysis of a system with parallel servers: Asymptotic optimality of discrete-review policies," *Annals of Applied Probability*, pp. 822–848, 1998.

[121] J. M. Harrison and M. J. López, "Heavy traffic resource pooling in parallel-server systems," *Queueing Systems*, vol. 33, no. 4, pp. 339–368, 1999.

[122] S. L. Bell, R. J. Williams et al., "Dynamic scheduling of a system with two parallel servers in heavy traffic with resource pooling: Asymptotic optimality of a threshold policy," *The Annals of Applied Probability*, vol. 11, no. 3, pp. 608–649, 2001.

[123] S. Bell, R. Williams et al., "Dynamic scheduling of a parallel server system in heavy traffic with complete resource pooling: Asymptotic optimality of a threshold policy," *Electronic Journal of Probability*, vol. 10, pp. 1044–1115, 2005.

[124] A. Mandelbaum and A. L. Stolyar, "Scheduling flexible servers with convex delay costs: Heavy-traffic optimality of the generalized cμ-rule," *Operations Research*, vol. 52, no. 6, pp. 836–855, 2004.

[125] T. Weller and B. Hajek, "Scheduling nonuniform traffic in a packet-switching system with small propagation delay," *IEEE/ACM transactions on networking*, vol. 5, no. 6, pp. 813–823, 1997.

[126] A. G. Greenberg and B. Hajek, "Deflection routing in hypercube networks," *IEEE Transactions on Communications*, vol. 40, no. 6, pp. 1070–1081, 1992.

[127] A. L. Stolyar, "Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic," *Annals of Applied Probability*, pp. 1–53, 2004.

[128] B. Hajek and R. G. Ogier, "Optimal dynamic routing in communication networks with continuous traffic," *Networks*, vol. 14, no. 3, pp. 457–487, 1984.

[129] M. Alanyali and B. Hajek, "Analysis of simple algorithms for dynamic load balancing," *Mathematics of Operations Research*, vol. 22, no. 4, pp. 840–871, 1997.

[130] B. Hajek and L. He, "On variations of queue response for inputs with the same mean and autocorrelation function," *IEEE/ACM Transactions on Networking*, vol. 6, no. 5, pp. 588–598, 1998.

[131] A. Yekkehkhany and R. Nagi, "Blind GB-PANDAS: A blind throughput-optimal load balancing algorithm for affinity scheduling," *IEEE/ACM Transactions on Networking*, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9042850

[132] W. Wang, K. Zhu, L. Ying, J. Tan, and L. Zhang, "Maptask scheduling in MapReduce with data locality: Throughput and heavy-traffic optimality," *IEEE/ACM Transactions on Networking*, vol. 24, no. 1, pp. 190–203, 2016.

[133] Q. Xie and Y. Lu, "Priority algorithm for near-data scheduling: Throughput and heavy-traffic optimality," in *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2015, pp. 963–972.

[134] Q. Xie, A. Yekkehkhany, and Y. Lu, "Scheduling with multi-level data locality: Throughput and heavy-traffic optimality," in *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on*. IEEE, 2016, pp. 1–9.

[135] S. Shakkottai and R. Srikant, "Scheduling real-time traffic with deadlines over a wireless channel," *Wireless Networks*, vol. 8, no. 1, pp. 13–26, 2002.

[136] B. Alinia, M. S. Talebi, M. H. Hajiesmaili, A. Yekkehkhany, and N. Crespi, "Competitive online scheduling algorithms with applications in deadline-constrained EV charging," in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*. IEEE, 2018, pp. 1–10.

[137] H. N. Matin, A. Yekkehkhany, and R. Nagi, "Probabilistic analysis of UAV routing with dynamically arriving targets," in *2019 22th International Conference on Information Fusion (FUSION)*. IEEE, 2019, pp. 1–8.

[138] M. Shafiee and J. Ghaderi, "An improved bound for minimizing the total weighted completion time of coflows in datacenters," *IEEE/ACM Transactions on Networking*, vol. 26, no. 4, pp. 1674–1687, 2018.

[139] A. Daw and J. Pender, "On the distributions of infinite server queues with batch arrivals," *Queueing Systems*, vol. 91, no. 3-4, pp. 367–401, 2019.

[140] E. Cardinaels, S. C. Borst, and J. S. van Leeuwaarden, "Job assignment in large-scale service systems with affinity relations," *Queueing Systems*, vol. 93, no. 3-4, pp. 227–268, 2019.

[141] T. Lambert, "On the effect of replication of input files on the efficiency and the robustness of a set of computations," Ph.D. dissertation, 2017.

[142] X. Zhou, J. Tan, and N. Shroff, "Flexible load balancing with multi-dimensional state-space collapse: Throughput and heavy-traffic delay optimality," *Performance Evaluation*, vol. 127, pp. 176–193, 2018.

[143] X. Zhou, F. Wu, J. Tan, K. Srinivasan, and N. Shroff, "Degree of queue imbalance: Overcoming the limitation of heavy-traffic delay optimality in load balancing systems," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 2, no. 1, pp. 1–41, 2018.

[144] X. Zhou, J. Tan, and N. Shroff, "Heavy-traffic delay optimality in pull-based load balancing systems: Necessary and sufficient conditions," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 2, no. 3, pp. 1–33, 2018.

[145] L. Bao, C. Q. Wu, H. Qi, W. Chen, X. Zhang, W. Han, W. Wei, E. Tail, H. Wang, J. Zhai et al., "Las: Logical-block affinity scheduling in big data analytics systems," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 522–530.

[146] X. Zhou, N. Shroff, and A. Wierman, "Asymptotically optimal load balancing in large-scale heterogeneous systems with multiple dispatchers," *arXiv preprint arXiv:2002.08908*, 2020.

[147] A. Budhiraja, "Theory and applications of weakly interacting markov processes," University of North Carolina-Chapel Hill Chapel Hill United States, Tech. Rep., 2018.

[148] Y. Raaijmakers, S. Borst, and O. Boxma, "Redundancy scheduling with scaled bernoulli service requirements," *Queueing Systems*, vol. 93, no. 1-2, pp. 67–82, 2019.

[149] G. Sanjay, G. Howard, and L. Shun-Tak, "The Google file system," in *Proceedings of the 17th ACM Symposium on Operating Systems Principles*, 2003, pp. 29–43.

[150] G. Ananthanarayanan, S. Agarwal, S. Kandula, A. Greenberg, I. Stoica, D. Harlan, and E. Harris, "Scarlett: Coping with skewed content popularity in MapReduce clusters," in *Proceedings of the Sixth Conference on Computer Systems*. ACM, 2011, pp. 287–300.

[151] C. L. Abad, Y. Lu, and R. H. Campbell, "Dare: Adaptive data replication for efficient cluster scheduling," in *Cluster Computing (CLUSTER), 2011 IEEE International Conference on*. IEEE, 2011, pp. 159–168.

[152] J. J. Jaramillo and R. Srikant, "Optimal scheduling for fair resource allocation in ad hoc networks with elastic and inelastic traffic," in *2010 Proceedings IEEE INFOCOM*. IEEE, 2010, pp. 1–9.

[153] S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," *IEEE/ACM Transactions on Networking*, vol. 7, no. 4, pp. 473–489, 1999.

[154] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," *IEEE/ACM Transactions on Networking*, vol. 15, no. 6, pp. 1333–1344, 2007.

[155] S. H. Lu and P. Kumar, "Distributed scheduling based on due dates and buffer priorities," *IEEE Transactions on Automatic Control*, vol. 36, no. 12, pp. 1406–1416, 1991.

[156] J. G. Dai and G. Weiss, "Stability and instability of fluid models for reentrant lines," *Mathematics of Operations Research*, vol. 21, no. 1, pp. 115–134, 1996.

[157] G. Baharian and T. Tezcan, "Stability analysis of parallel server systems under longest queue first," *Mathematical Methods of Operations Research*, vol. 74, no. 2, p. 257, 2011.

[158] A. Dimakis and J. Walrand, "Sufficient conditions for stability of longest-queue-first scheduling: Second-order properties using fluid limits," *Advances in Applied probability*, vol. 38, no. 2, pp. 505–521, 2006.

[159] R. Pedarsani and J. Walrand, "Stability of multiclass queueing networks under longest-queue and longest-dominating-queue scheduling," *Journal of Applied Probability*, vol. 53, no. 2, pp. 421–433, 2016.

[160] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[161] M. Isard, M. Budiu, Y. Yu, A. Birrell, and D. Fetterly, "Dryad: Distributed data-parallel programs from sequential building blocks," in *ACM SIGOPS operating systems review*, vol. 41, no. 3. ACM, 2007, pp. 59–72.

[162] S. Kavulya, J. Tan, R. Gandhi, and P. Narasimhan, "An analysis of traces from a production MapReduce cluster," in *Cluster, Cloud and Grid Computing (CCGrid), 2010 10th IEEE/ACM International Conference on*. IEEE, 2010, pp. 94–103.

[163] Y. Chen, S. Alspaugh, D. Borthakur, and R. Katz, "Energy efficiency for large-scale MapReduce workloads with significant interactive analysis," in *Proceedings of the 7th ACM European Conference on Computer Systems*. ACM, 2012, pp. 43–56.

[164] G. Ananthanarayanan, A. Ghodsi, A. Wang, D. Borthakur, S. Kandula, S. Shenker, and I. Stoica, "Pacman: Coordinated memory caching for parallel jobs," in *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*. USENIX Association, 2012.

[165] Q. Xie, M. Pundir, Y. Lu, C. L. Abad, and R. H. Campbell, "Pandas: Robust locality-aware scheduling with stochastic delay optimality," *IEEE/ACM Transactions on Networking*, 2016.

[166] J. Vermorel and M. Mohri, "Multi-armed bandit algorithms and empirical evaluation," in *European Conference on Machine Learning*. Springer, 2005, pp. 437–448.

[167] H. Robbins, "Some aspects of the sequential design of experiments," *Bulletin of the American Mathematical Society*, vol. 58, no. 5, pp. 527–535, 1952.

[168] D. Bergemann and U. Hege, "Dynamic venture capital financing, learning and moral hazard," *Journal of Banking and Finance*, vol. 22, no. 6-8, pp. 703–735, 1998.

[169] D. Bergemann and U. Hege, "The financing of innovation: Learning and stopping," *RAND Journal of Economics*, pp. 719–752, 2005.

[170] O. Avner and S. Mannor, "Multi-user lax communications: A multi-armed bandit approach," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[171] D.-S. Zois, "Sequential decision-making in healthcare IoT: Real-time health monitoring, treatments and interventions," in *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*. IEEE, 2016, pp. 24–29.

[172] N. Musavi, D. Onural, K. Gunes, and Y. Yildiz, "Unmanned aircraft systems airspace integration: A game theoretical framework for concept evaluations," *Journal of Guidance, Control, and Dynamics*, pp. 96–109, 2016.

[173] R. Meshram, D. Manjunath, and A. Gopalan, "On the Whittle index for restless multiarmed hidden Markov bandits," *IEEE Transactions on Automatic Control*, vol. 63, no. 9, pp. 3046–3053, 2018.

[174] S. Maghsudi and E. Hossain, "Distributed user association in energy harvesting dense small cell networks: A mean-field multi-armed bandit approach," *IEEE Access*, vol. 5, pp. 3513–3523, 2017.

[175] A. Lesage-Landry and J. A. Taylor, "The multi-armed bandit with stochastic plays," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 2280–2286, 2017.

[176] A. Yekkehkhany, E. Arian, M. Hajiesmaili, and R. Nagi, "Risk-averse explore-then-commit algorithms for finite-time bandits," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019.

[177] N. V. Chawla and D. A. Davis, "Bringing big data to personalized healthcare: A patient-centered framework," *Journal of General Internal Medicine*, vol. 28, no. 3, pp. 660–665, 2013.

[178] D. E. Pritchard, F. Moeckel, M. S. Villa, L. T. Housman, C. A. McCarty, and H. L. McLeod, "Strategies for integrating personalized medicine into healthcare practice," *Personalized Medicine*, vol. 14, no. 2, pp. 141–152, 2017.

[179] K. Priyanka and N. Kulennavar, "A survey on big data analytics in health care," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5865–5868, 2014.

[180] E. Abrahams, G. S. Ginsburg, and M. Silver, "The personalized medicine coalition," *American Journal of Pharmacogenomics*, vol. 5, no. 6, pp. 345–355, 2005.

[181] A. Garivier, P. Ménard, and G. Stoltz, "Explore first, exploit next: The true shape of regret in bandit problems," *Mathematics of Operations Research*, 2018.

[182] L. Prashanth, "Cs6046: Multi-armed bandits," *course notes, IIT Madras*, 2018.

[183] T. Lattimore and C. Szepesvári, *Bandit algorithms*. Cambridge University Press, 2020.

[184] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[185] M. Swider, "$3bn up in smoke: Samsung reveals losses on Galaxy note 7 recall," *Techradar*, Oct 2016. [Online]. Available: https://www.techradar.com/news/this-is-how-much-samsung-says-itll-lose-on-the-galaxy-note-7-recall

[186] J. McCallion, "Amazon loses $170 million on fire phone," *Alphr*, Oct 2014. [Online]. Available: https://www.alphr.com/news/391360/amazon-loses-170-million-on-fire-phone

[187] T. Warren, "Microsoft wasted at least $8 billion on its failed Nokia experiment," *The Verge*, May 2016. [Online]. Available: https://www.theverge.com/2016/5/25/11766540/microsoft-nokia-acquisition-costs

[188] A. Yekkehkhany, E. Arian, R. Nagi, and I. Shomorony, "A cost-based analysis for risk-averse explore-then-commit finite-time bandits."

[189] J. G. Wardrop, "Some theoretical aspects of road traffic research," *Proceedings of the Institution of Civil Engineers*, vol. 1, no. 3, pp. 325–362, 1952.

[190] M. Abdel-Aty, R. Kitamura, and P. P. Jovanis, "Investigating effect of travel time variability on route choice using repeated-measurement stated preference data," *Transportation Research Record*, no. 1493, 1995.

[191] C. Kazimi, A. Brownstone, and T. Gosh, "Willingness-to-pay to reduce commute time and its variance: Evidence from the San Diego I-15 congestion pricing project," in *Transportation Research Board 79th Annual Meeting [CD-ROM]*, Washington, DC, 2000.

[192] T. Lam, "The effect of variability of travel time on route and time-of-day choice," Ph.D. dissertation, Department of Economics, University of California, Irvine, Claif, USA, 2000.

[193] T. C. Lam and K. A. Small, "The value of time and reliability: Measurement from a value pricing experiment," *Transportation Research Part E: Logistics and Transportation Review*, vol. 37, no. 2-3, pp. 231–251, 2001.

[194] K. A. Small et al., *Valuation of Travel-time Savings and Predictability in Congested Conditions for Highway User-cost Estimation*. Transportation Research Board, 1999.

[195] D. Braess, "Über ein paradoxon aus der verkehrsplanung," *Unternehmensforschung*, vol. 12, no. 1, pp. 258–268, 1968.

[196] J. D. Murchland, "Braess's paradox of traffic flow," *Transportation Research*, vol. 4, pp. 391–394, 1970.

[197] A. C. Pigou, *The Economics of Welfare*. Macmillan and Company, Limited, 1920.

[198] A. C. Pigou, *The Economics of Welfare*. Palgrave Macmillan, 2013.

[199] D. Padua, *Encyclopedia of Parallel Computing*. Springer Science & Business Media, 2011.

[200] A. Yekkehkhany, A. Hojjati, and M. H. Hajiesmaili, "GB-PANDAS: Throughput and heavy-traffic optimality analysis for affinity scheduling," *ACM SIGMETRICS Performance Evaluation Review*, vol. 45, no. 2, pp. 2–14, 2018. [Online]. Available: https://doi.org/10.1145/3199524.3199528

[201] A. Yekkehkhany, "Near-data scheduling for data centers with multiple levels of data locality," M.S. thesis, University of Illinois at Urbana-Champaign, 2017.

[202] R. Pedarsani, J. Walrand, and Y. Zhong, "Robust scheduling for flexible processing networks," *Advances in Applied Probability*, vol. 49, no. 2, pp. 603–628, 2017.

[203] R. Srikant and L. Ying, *Communication Networks: An Optimization, Control, and Stochastic Networks Perspective*. Cambridge University Press, 2013.

[204] Q. Xie, "Scheduling and resource allocation for clouds: Novel algorithms, state space collapse and decay of tails," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2016.

[205] A. Downs, "An economic theory of political action in a democracy," *Journal of Political Economy*, vol. 65, no. 2, pp. 135–150, 1957. [Online]. Available: http://www.jstor.org/stable/1827369

[206] D. Mukherjee, S. C. Borst, J. S. van Leeuwaarden, and P. A. Whiting, "Universality of power-of-d load balancing in many-server systems," *arXiv preprint arXiv:1612.00723*, 2016.

[207] Y. Lu, Q. Xie, G. Kliot, A. Geller, J. R. Larus, and A. Greenberg, "Join-idle-queue: A novel load balancing algorithm for dynamically scalable web services," *Performance Evaluation*, vol. 68, no. 11, pp. 1056–1071, 2011.