Regulating Self-Adaptive Multi-Agent Systems with Real-Time Interventions

by

Wen Shen

A Thesis Presented to the

Masdar Institute of Science and Technology

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in

Computing and Information Science

Regulating Self-Adaptive Multi-Agent Systems with Real-Time Interventions

By Wen Shen

A Thesis Presented to the Masdar Institute of Science and Technology in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computing and Information Science

June 2013

AUTHOR'S DECLARATION

I understand that copyright in my thesis is transferred to Masdar Institute of Science and Technology.

Author ______________________

RESEARCH SUPERVISORY COMMITTEE MEMBERS

Dr. Jacob Crandall, Chair, ______________________
Masdar Institute of Science and Technology

Dr. Iyad Rahwan, ______________________
Masdar Institute of Science and Technology

Dr. Zeyar Aung, ______________________
Masdar Institute of Science and Technology

Abstract

Dynamic resource allocation in multi-agent systems has numerous critical applications, especially in solving some very challenging problems related to sustainability. These applications include traffic management in transportation systems, energy management in building systems and demand side management in power grids and water supply network.

Since resources in these systems are bounded or limited, agents need to share and compete for the limited resources with peers. The agents are autonomous or self-motivated. They are not directly controlled by the regulatory entities, though they are subject to regulations implemented by the regulators. However, the preferences of the agents are subject to change and not fully available to the regulators due to privacy or other considerations. These systems may not meet societal goals without proper regulations or management. Therefore, it is beneficial to study the principles of interventions in these systems.

In this work, we investigate whether, under what conditions and to what extent people are able to create effective interventions to manage these multi-agent systems in real-time. We present an abstract transportation framework named *Jiao Tong* to model the dynamic resource allocation problems. We then develop a user interface based on Jiao Tong, which enables users to create real-time interventions. We propose six hypotheses with respect to Jiao Tong and then perform and analyze user studies to test these hypotheses.

This research was supported by the Government of Abu Dhabi to help fulfill the vision of the late President Sheikh Zayed Bin Sultan Al Nayhan for sustainable development and empowerment of the UAE and humankind.

Acknowledgments

I would like to express my sincerest gratitude and appreciation to my advisor, Dr. Jacob W. Crandall, for his guidance and continual support throughout my studies at Masdar Institute. His encouragement, patience and kindness inspired me and led me to the right way in my academic pursuit.

I am grateful to my thesis committee members, Dr. Iyad Rahwan and Dr. Zeyar Aung, for their reviews and constructive advice on this thesis. I would like to thank Dr. Wei Lee Woon for his help in my courses as well as research.

I would like to extend my appreciation to the volunteers who participated in the human subject experiments in my study for their precious time and helpful suggestions. I also want to thank my lab mates, Abdulla Almehrezi, Alanoud Al Khemiri, Issak Gezehei, Rafael Harutyunyan, Olga Tsvetkova, Vahagan Harutyunyan, Vimitha Monohar and Waleed K.A Najy for their help and the joy that they brought. I am indebted to my other friends at Masdar Institute and beyond, though I am unable to list their names completely. My life would not have been so joyful without them.

My deepest appreciation goes to my family for their abiding love, understanding and support. This thesis is dedicated to them.

Contents

List of Figures

CHAPTER 1

Introduction

In *multi-agent systems* (MAS) such as *building management systems* (BMS), water supply systems, power grids and transportation systems, the phenomenon of sharing and competing for limited resources among multiple autonomous, self-interested and adapting agents (human or artificial) is pervasive. In these systems, the regulatory entities or the supervisory departments (e.g., the building operators, water distribution operators, independent electricity system operators, or the transportation authorities) are motivated to balance the needs of different agents or entities, minimize the consumption of energy or resources and reduce the occurrences of traffic congestion.

However, agents may adapt to satisfy their own interests and preferences and their aggregate behavior may not conform to societal goals. For example, during rush hours, drivers select routes based on their own needs and wants which may not be known to others. If the number of vehicles that are simultaneously running on a road is at or near the full capacity of the road, then traffic congestion or a traffic jam will be likely to occur. Similar situations (e.g., power outages due to insufficient supply during peak demand periods in electric power industry) also happen in other resource-bounded and self-adaptive MAS. In order to fulfill the societal goals and maximize the benefits of the whole society, the regulators need to take measures to promote desirable behaviors and deter undesirable behaviors.

1.1 Motivations and Objectives

To build a sustainable society, it is necessary for decision makers to promote environmentally friendly (or eco-friendly) behaviors and encourage people to utilize resources efficiently and wisely. Humans are self-interested and are able to continually learn from past experience and adapt to the dynamic environment. Without necessary management or supervision, they may act solely based on their individual needs and neglect to consider the interests of others or the whole society, which may hamper the process of building an environmentally and ecologically sustainable world. Therefore, it is important and beneficial to study the principles and mechanisms of interventions in resource-bounded and self-adpative MAS.

In this work, we try to better understand the general principles of real-time interventions for efficient management of self-adaptive, resource-limited MAS. We focus on identifying what aspects of these systems allow regulators to effectively manage the consumption of valuable, potentially scarce resources. In particular, we seek to investigate whether, under what conditions and to what extent people are able to create effective interventions in real-time to solve *dynamic resource allocation problems* (DRAPs) in resource-bounded and self-adaptive MAS.

To determine effective interventions for these systems, several aspects of problems need to be considered. First, since the preferences and adaptation patterns of individual agents are various and not fully known to regulators, it is difficult and expensive to model the dynamics of each individual agent's adaptation by using traditional multi-agent learning algorithms such as reinforcement learning. However, it is necessary for the regulators to perceive the behavioral patterns of the agents in order to make wise decisions on interventions. Hence, modeling the dynamics of agents' behavior from the perspective of the regulators becomes a challenging problem to deal with.

Second, it is also demanding for regulators to determine the right time to take actions when the systems are in undesired states. For instance, in a transportation network, if the traffic flow on a specific road is increasing over time, the probability of a potential traffic jam on the road might also increase. To maximize the throughput of the network, regulators need to decide when to initiate interventions. If they continue to collect data of the state, they may miss the best time to intervene. If they take actions too early, they may limit the reasonable growth of the throughput. Therefore, it is important to determine the right time to start interventions.

Third, an intervention method that would be effective in one situation might not work in another scenario because of the adaptation of the agents and the changing environment. In order to make wise decisions, regulators must be able to identify the available intervention methods as well as the applicable situations of these methods.

Fourth, it is not robust to fully rely on humans to make the decisions on interventions because humans are prone to suffer cognitive and other biases that might lead to bad judgments and poor decision-making [27]. Moreover, factors such as fatigue can produce random fluctuations in judgment [11]. Automated control systems may also be unreliable because of failures of automated devices. Thus, it is important to design a robust and efficient decision-making framework for regulators.

1.2 Main Contributions

This work is designed to make several contributions:

1. We design an abstract DRAP framework, which we call the *Jiao Tong* [1] game. The game has attributes of many resource allocation problems in real-world applications such as transportation systems, BMS, water distribution systems and smart grids.

2. We develop an interactive simulator based on Jiao Tong, which enables users (act as regulators) to create real-time interventions by giving tolls on each route.

3. We conduct two user studies to investigate whether, under what conditions and to what extent people are able to effectively create interventions to manage resource-limited and self-adaptive MAS in real-time. We also evaluate people's capabilities to perceive and learn agents' aggregate behaviors.

1.3 Thesis Organization

The thesis is organized as follows: Chapter 1 presents the background, motivation and objectives of this work, followed by the main contributions and the organization of the thesis. Chapter 2 discusses the related work on approaches to solving the DRAPs, and methodologies to model and simulate the agents in the MAS. Chapter 3 formally defines the Jiao Tong game. Chapter 4 and Chapter 5 present two user studies in which participants seek to manage autonomous agents using real-time interventions in Jiao Tong. Chapter 6 summarizes the main contributions of the thesis and discusses future work.

[1] Jiao Tong means *traffic, transportation and communication* in Chinese.

CHAPTER 2

Related Work

The main objective for a regulatory entity managing a resource-bounded MAS is to wisely allocate the limited resources and properly coordinate all agents' necessity so that societal goals can be achieved. In this chapter, we start by introducing related work on resource allocation from several aspects. We then describe the models used to simulate agents in resource-limited MAS, followed by a discussion on intervention strategies.

2.1 Mechanism Design for Resource Allocation

One of the key challenges in regulating multi-agents systems is to accurately and promptly learn the behavioral patterns or the internal dynamics of agents. Mechanism design [48, 36, 29] has been intensively studied and widely applied to resource allocation domains such as task scheduling [37, 22, 31], load balancing [36, 19] and routing [36, 21, 14]. The general idea of mechanism design in resource allocation domains is to design dominant incentive compatible mechanisms which induce agents or people to elicit their true preferences or decision-making patterns by offering monetary or other kinds of incentives. That is, each agent receives its maximum utility by truthfully revealing its genuine preferences to the regulatory entities. These truth revelations become a dominant strategy for each agent in the MAS. Based on the private information (or the preferences) obtained from the agents, the regulatory entities (or the system managers) figure out the optimal allocation solutions and make final decisions to maximize the expected utilities from the perspective of the society.

Recent advances in smart grids and electric vehicle charging show that mechanism design is gaining favor. Rose et al. [39, 40] proposed a spherical scoring rule-based mechanism named *sum of others' plus max* to collect the truthful estimation data of agents' energy consumption for the aggregator by providing incentives. Stein et al. [43] presented an online mechanism based on the Consensus algorithm [4] to schedule the charging of electric vehicles. They argued both theoretically and empirically that these mechanisms worked well in the respective domains. The prerequisite for this kind of mechanism design is that a priori knowledge of the agents' internal dynamics (e.g. the agents' state and action spaces) is available.

However, in many real-world applications, a prior knowledge of such kind is usually unknown or not available. For example, due to the high interdependency and interconnectivity of the transportation network, a tiny unpredictable accident on a road may change the traffic situation of a district significantly and exert a large impact on the whole city network. New phenomena may emerge and the decision spaces for agents may alter. Moreover, drivers may not be willing to disclose their preferences because of privacy or other considerations even if monetary incentives for truthfully revealing preferences are provided. In such cases, solving the highly

DRAPs by simply employing traditional approaches to mechanism design becomes extremely difficult.

2.2 Approximation Algorithms for Resource Allocation

There also has been a lot of interests recently in approximation algorithms [24, 23, 30, 49] for resource allocation. Saha [41] presented a detailed investigation on approximation algorithms and their applications to resource allocation problems such as scheduling and fair allocation. They introduced a rounding approximation algorithm which incorporated the concepts of dependent rounding and iterative relaxations to solve the resource allocation problems. Calinescu et al. [6] proposed several approximation algorithms including the Large Tasks Algorithm, the Small Tasks Algorithm and the List Algorithm to deal with the pre-defined resource allocation problem where some renewable resource and time were consumed for producing profits. Chuzhoy et al. [7] presented a polynomial approximation algorithm for job interval selection problems which could produce a better approximation guarantee.

AuYong et al. [3] studied the applications of different approximation algorithms to allocation problems in distributed computing systems. These applications belong to the combinatorial optimization domain and most of them have been shown to be NP-hard. The conditions of these problems are clear and known to the system designers or managers. Therefore, approximation algorithms may work well if elegantly designed. Nevertheless, in highly dynamic multi-agent systems, this assumption may not always hold because the environment may be subject to change dramatically. Without good understanding of the system dynamics, it is extremely difficult to design effective and efficient approximation algorithms for solving such DRAPs.

2.3 Machine Learning for Resource Allocation

Machine learning techniques also have great potential for solving resource allocation problems. Cs*á*ji et al. [8, 9] represented the transition function by using support vector regression and hash tables, and employed reinforcement learning methods such as Markov decision processes and approximate dynamic programming to manage adaptive and stochastic resource allocation problems. Galstyan et al. [15, 16] and Li et al. [33] studied reinforcement learning algorithms in the allocation of grid computing resources. Tesauro [45] introduced a decompositional reinforcement learning method for an online resource allocation problem where servers were allocated in real time in order to maximize the overall expected utilities. These machine learning methods were proved by the investigators involved to work well in the respective domains because they could perceive the dynamic environment through learning. In spite of that, most current learning algorithms scale poorly when learning increasingly complex concepts and environments due to the fast changing environment, partial observation of the states and unknown decision space of the agents. For example, machine learning techniques typically require vast amounts of examples to distinguish the relevant from the irrelevant, while humans are often able to generalize correctly and acquire knowledge from very small training dataset, even with a large number of potentially relevant features [46].

2.4 Human Supervisory Control for Resource Allocation

Recent work in the area of collective intelligence [47, 51] shows that it is more accurate and robust to combine human and machine intelligence than only use either type alone to make predictions in situations where patterns are difficult to discern [35]. Kamar et al. [28] proposed a system framework that combines the efforts of people and machine learning to make effective decisions on consensus tasks such as when to hire workers and how to perform classifications when observations cease. They demonstrated that such methodology could solve consensus tasks accurately and save resources significantly.

In highly dynamic and complex systems such as smart buildings, transportation networks and power girds, it is difficult or not feasible to predict agents' future behavioral patterns either manually or automatically. The reasons are described as follows: first, regulatory entities may not have a prior knowledge of the preferences or internal dynamics of the agents; second, the state and action spaces for agents may be unknown and subject to change, so the dimension of decision spaces for the agents is probably not available to the regulators. Therefore, compromises between human supervision and automation are required in the design of these control systems. In our work, we propose a framework which enables regulators to design real-time interventions.

Human Supervisory Control [42] has been intensively studied in the design of automated systems consisting of autonomous agents, e.g., mobile robot systems [1] and unmanned aerial vehicles [10]. In human supervisory control of these dynamic and highly automated systems, a human operator is in charge of observing the system state, detecting anomalies, collecting data and taking actions to bring the systems back to normal states [26]. In these systems, the agents are controlled by the human operators directly or indirectly through the commands given by human operators. In scenarios like building management systems and transportation networks, however, the preferences of individuals are not necessarily aligned with societal (or the regulators') goals. Autonomous agents in these systems are not fully controlled by regulators.

2.5 Modeling The Dynamics of Agents

Understanding the nature of group behavior [34] in situations in which limited resources must be shared is necessary prior to the development of intervention systems. Agents' behavior can be generally divided into two types: rational behavior and irrational behavior. Rational behavior means that an agent will always respond based on factual information in order to maximize its interests or the social benefits. Irrational behavior means that an agent will behave in a way that is neutral or detrimental to its interests [17]. In order to determine the efficiency of intervention strategies, it is necessary to model the dynamics of individual agents in the system. However, from the perspective of regulators, it is expensive or even unfeasible to learn each agent's true preferences. Nevertheless, it is possible to build computational models of the collective behavior of agent groups [44] and relatively inexpensive to make predictions based on these models using statistical methods. For example, to predict the traffic situation on a road, regulators may only need to learn the trends of the traffic flows instead of the preferences of each driver.

Human behavior is very difficult to illustrate or model because of the unknown processes of cognition. However, research has shown that there exist some basic laws that the majority of people will follow in complex systems like transportation systems [20]. In their work, Helbing et al. argued that individuals normally behave rationally in reality, which means avoiding traffic jams and delays, minimizing fuel costs, and maximizing external benefits if applicable. In conclusion, agents are self-interested and self-motivated. In our work, we consider two agent models: learning[1] and static[2]. Agents in both the learning model and the static model are self-interested. That is, they make decisions to maximize their utilities.

[1] Learning model means all the agents are learning (or adapting) agents which adapt to the environment.

[2] Static model means all the agents are static (or non-adapting) agents which do not adapt to the environment.

The reason we select these two models is that rational agents (human or artificial) may either adapt to the changing environment or do not adapt to it. For example, consider a road where congestion fees are charged when the traffic demand exceeds its capacity. Drivers who have other alternative routes to their destinations may try to avoid this road if they estimate that there will be congestion based on their experience. However, drivers who do not have other choices may have to select the road even if they know that there will be jams on the road. Different learning agents may have different learning capabilities. These capabilities are subject to change over time. However, all the learning agents are able to accumulate experience and knowledge of the environment through perception and interaction. Static agents do not learn, but they may respond based on available information or knowledge.

2.6 Intervention Strategies

A common method for group control is via explicit rules or laws [18]. We call such laws *Force Interventions*. For instance, if we wish to relieve the traffic pressure on a road, we may simply close the road.

Individual and group behavior can also be regulated by providing monetary rewards or other material incentives to agents who contribute to societal goals or applying penalties to agents who perform undesired behaviors. These methods are defined as *Incentive Interventions*. One popular application of incentive interventions is real-time pricing in power grids [5, 38].

In systems such as smart buildings or transportation networks, agents only have a local view while the regulator has a partial global view of the system state. Therefore, it is possible for the regulator to mold the behavior of agents by providing certain global information to them. These methods belong to *Information Inter-*

ventions.

In our work, we only focus on incentive interventions. The reasons are described as follows. First, incentive interventions are widely used in real world. For example, congestion charging on vehicle emissions is considered to be a common intervention method. Other practices include smart meters in power grids and water supply network. Second, it is measurable. Since incentives or tolls are provided in order to make interventions, it is convenient to evaluate the impact or the effects of interventions both quantitatively and qualitatively.

CHAPTER 3

The Jiao Tong Game

In this chapter, we describe the Jiao Tong game. We also introduce two different models, the learning (or adapting) model and the static (or non-adapting) model, used to simulate agents' behaviors.

3.1 The Definition

Consider a general network $G = (N,E)$, where N denotes the set of nodes (vertices), and E the set of links (edges). Let $N = \{A, B, C, D\}$, $E = \{\vec{AB}, \vec{AC}, \vec{BC}, \vec{CB}, \vec{BD}, \vec{CD}, \vec{DA}\}$. The network G (as shown in Fig. 3.1) is a directed graph which

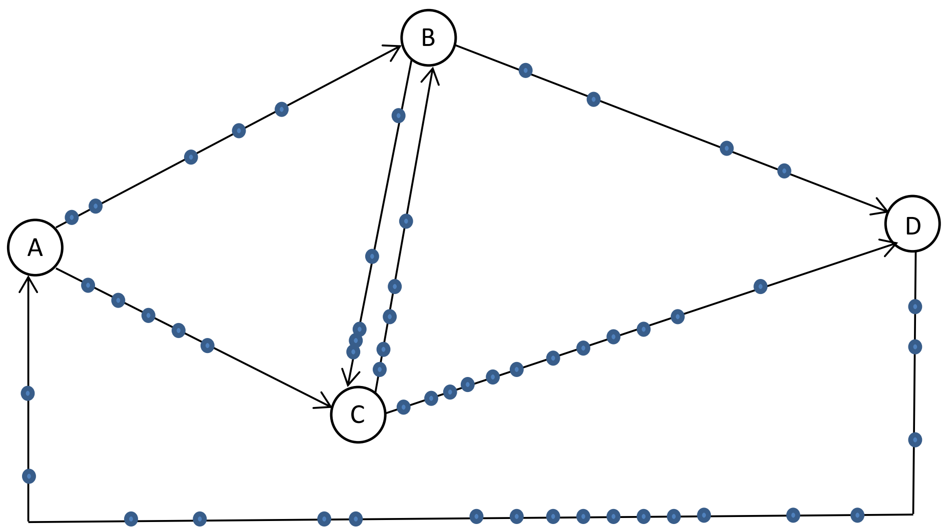

Figure 3.1: The network scenario: A, B, C, D are the 4 nodes in set N; AB, AC, BC, BD, CB, CD and DA are the 7 one-way links in set E; the solid circles on the links represent the self-interested, autonomous agents moving across the network.

consists of 4 nodes and 7 links. It can be represented by the following matrix:

$$G = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array}, \tag{3.1}$$

where "1" means a directed link exists in the network and "0" indicates the non-existence of a link. The coordinate of node i is represented as: (x_i, y_i). The distance between node i and node j is calculated by the following equation:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{3.2}$$

Each link ij has a capacity Cap_{ij} which is the maximum traffic flow or the maximum number of agents (vehicles) that can move at a relatively high velocity compared to the maximum speed allowed. If the number of agents on a link nears or exceeds the capacity, then the traffic speed on this link will slow down according certain rule described later in this section. Let Num_{ij}^{t} denote the number of agents on link ij at time t, V_{ij} be the maximum speed of the agent on link ij, let σ represent a random number ranging from 0 to 1, and let R_1 and R_2 be constants where $R_1 > R_2$. In the experiment, $R_1 = 10$ and $R_2 = 5$. The velocity of agent m on link ij at time t is defined as follows:

$$v_{ij}^{m} = (V_{ij} + \delta_m) \cdot f_{ij}^{t}, \tag{3.3}$$

where δ_m is an adjustment of the velocity following a uniform distribution:

$$\delta_m = \sigma \cdot R_1 - R_2, \tag{3.4}$$

and f_{ij}^{t} is an load factor determined as following:

$$f_{ij}^{t} = 0.9 \cdot \left(\frac{1}{1 + e^{0.25(Num_{ij}^{t} - Cap_{ij})}}\right) + 0.1. \tag{3.5}$$

As shown in Fig. 3.2, the load factor decreases gradually with the number of agents on the link increasing from 0 to half of the link capacity. As the number of agents raises, the load factor first drops dramatically and followed by a slow fall approaching 0.1 infinitely.

There are M agents in this network. Initially, the M agents are randomly allocated to the four nodes. Then, the agents move across the network based on their individual preferences which will be discussed in Section 3.3.

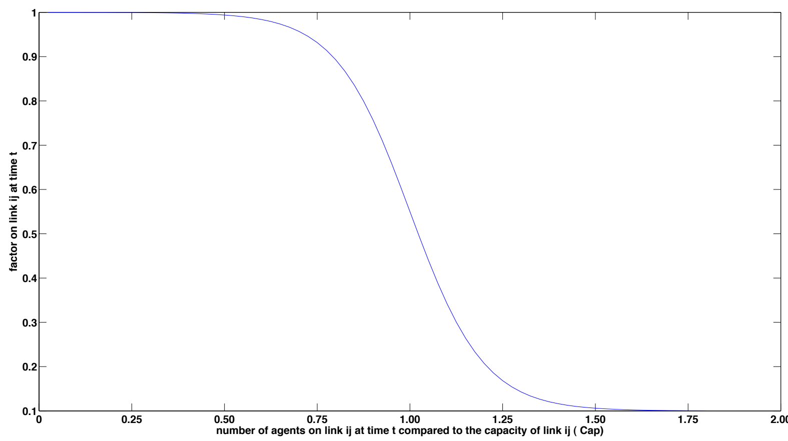

Figure 3.2: The load factor determined by the number of agents on the link and the capacity of that link.

3.2 The Regulator

In Jiao Tong, the objective of a regulator is to maximize the *running throughput* of the network by giving tolls on the links where interventions are available. There are six links which enable regulators to initiate toll controls: $\vec{AB}$, $\vec{AC}$, $\vec{BC}$, $\vec{BD}$, $\vec{CB}$ and $\vec{CD}$. To support regulators' decision making, we introduce another criteria, *sliding throughput*, which reflects the most recent states of the network. Both metrics are defined as follows:

> ***Running Throughput***: Running throughput is the average number of agents (vehicles) that have passed through node D per second since the game starts.

> ***Sliding Throughput***: Sliding throughput is the average number of agents (vehicles) that have passed through node D per second over the last 30 seconds.

Regulators are expected to observe the traffic status and the aggregate behavior of agents in the network. Based on the traffic situation on each link[1] and the movement of the crowds, regulators determine the tolls on each link. Regulators may charge relatively high fees to agents who enter the congested links and charge less to those who choose the links with less traffic. For example, if link $\vec{CD}$ is overcrowded, a regulator may have the following options to reduce the traffic congestion: increase the toll on link $\vec{CD}$, decrease the road pricing on link $\vec{CB}$, increase the surcharge on link $\vec{AC}$, decrease the fees on link $\vec{AB}$ or combinations of these methods. By properly balancing the tolls on each link, regulators can induce the traffic flow to quickly pass through node D so that high running throughput can be achieved.

In Jiao Tong, regulators are allowed to increase or decrease tolls on each link by \$0.01 (left click) and \$0.05 (right click) each time. The tolls on each link could be ranged from \$0.00 to \$0.99. Initially, tolls are set to \$0.50 for all the link in set $E^{'}$.

Before the game starts, the regulator is informed that: agents in this game may differ in selecting nodes to visit and in choosing links to traverse; all the agents have full information of tolls on each link; they use tolls to help them make their decisions. The decision-making process of agents in selecting nodes and links is initially unknown to the regulator. However, it is possible for the regulator to learn it from agents' aggregate behavior through interaction and observation. During the game, the regulator may choose to observe the performance graph of running throughput as well as the dynamic curve of sliding throughput to support their decision making. The status of each link, including the capacity of the link and the current number of agents running on the link, are also available to the regulator. The regulator can evaluate the interventions he/she has created by observing the

[1]Unless explicitly defined, *link* occurred in the remaining part of this thesis refers to a link which enables regulators to control it through tolls. In other words, a link that is in set $E^{'}$.

sliding throughput curve. If the sliding throughput increases, it indicates that the interventions are effective. Otherwise, the regulator needs to rethink his/her control strategies.

In the learning model, agents learn the time cost for a traversal of each link from their experience. Once the network has achieved a congestion-free state (or a dynamic equilibria), it needs little or no effort for regulators to maintain the state. In the static model, however, agents do not learn the recent traffic situation, even if a dynamic equilibria is obtained, intervention is still necessary. In this work, we investigate people's behavior in regulating self-interested, autonomous agents in Jiao Tong by creating real-time interventions through experimental analysis.

To measure the efforts a regulator has made in the real-time interventions, we define another metric-*toll changes*. Let ψ be the duration of the game, T_{ij}^{t} denote the toll on link ij at time t (the time is counted from 0), Φ be the toll changes and $E^{'}$ be the set of links which could be controlled by regulators through tolls. Here, $E' \subset E$ and $E' = \{\vec{AB}, \vec{AC}, \vec{BC}, \vec{BD}, \vec{CB}, \vec{CD}\}$. We have:

$$\Phi = \sum_{ij \in E} \int_{0}^{\Psi} |T_{ij}^{t+1} - T_{ij}^{t}| \, \mathrm{d}t \tag{3.6}$$

3.3 Models for Simulating The Agents

In this section, we formally define the two models used to simulate the agents' behavior in this thesis: the learning model and the static model. In Jiao Tong, the decision making process of agents consists of three phases: selecting a goal node, estimating time to traverse the links and selecting a link to traverse. Whenever an agent (learning or static) arrives at a node, it selects a goal node based on its preference. It then estimates the time cost for traversing each link and chooses the path with minimum cost to pass over. The general process is shown in Fig. 3.3.

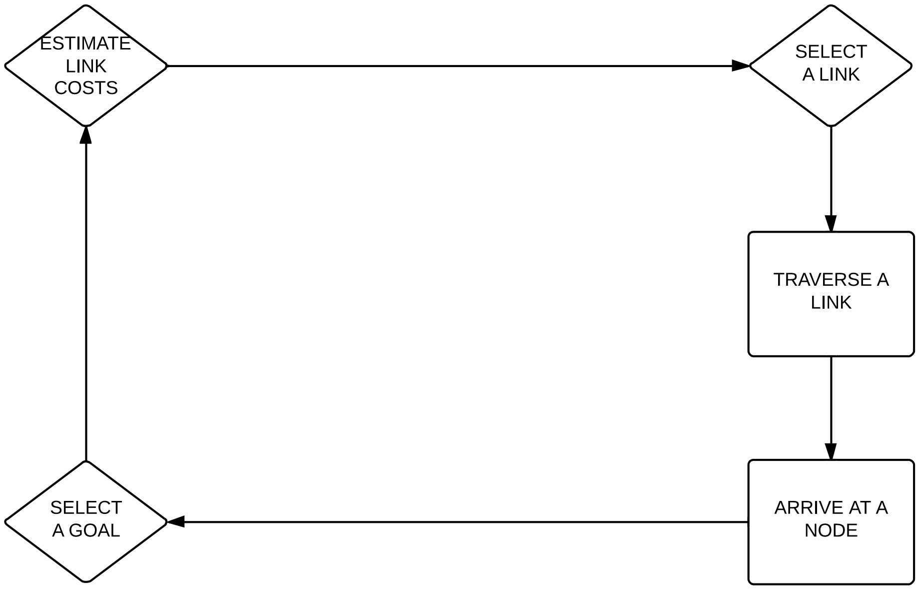

Figure 3.3: The process of agents' decision making in Jiao Tong.

3.3.1 The Learning Model

A learning agent updates its estimate of time costs each time it finishes a traversal of a link. After arriving at a node, it receives its utility, selects the goal node which could bring the maximum utility and chooses the link with minimum cost to its goal. A learning agent learns the amount of time it takes to traverse each link from its experience.

Selecting a goal

Each time an agent m reaches a node (whether the node is its current goal or some other node), the agent re-selects a goal $g \in N$. It adheres to the three steps described as follows:

1. Estimate the cost of going to each node $g \in N$ from node i. Let D^t_{ig} denote this expected cost. D^t_{ig} is computed using Dijkstra's algorithm [12, 25] based

on the following equation:

$$D_{ig}^{t} = \max_{j \in N_i} \left(C_{ij}^{t} \theta + T_{ij}^{t} + G_{jg}^{t} \right), \tag{3.7}$$

where $N_i \in N$ is the set of nodes that are adjacent to node $i \in N$, C_{ij}^t is the agent's estimate at time t of how long it will take for it traverse link ij, θ is the agent's operation cost per unit time, T_{ij}^t is the toll at time t for entering link ij, and G_{ij}^t is the agent's estimate utility for choosing goal node $g \in N$ when it is at node $j \in N$ at time t.

2. Compute $G_{ig}^t = U_i^t - D_{ig}^t$ for each $g \in N$, where U_i^t is the utility that the agent places on getting to node $i \in N$ at time t. Each time the agent reaches its currently selected goal, U_i^t for each $i \in N$ is re-established by drawing a sample from a normal distribution. The distributions are selected uniquely for each node and each agent at the beginning of the game.

3. Select $g^* = \max_{g \in N} G_{ig}^t$ as the current goal location.

An agent only gets utility for arriving at a goal node. If node i is selected as the goal node, the agent must leave node i before coming back to it. An agent will be only charged when it enters a link where the toll on the link is greater than zero.

Estimating time to traverse a link

Initially, the estimated time cost C_{ij}^t for a traversal of link ij by a learning agent:

$$C_{ij}^{0} = \frac{d_{ij}}{V_{ij}}, \tag{3.8}$$

where d_{ij} and V_{ij} are defined in Equation 3.2, 3.3, respectively. Each time the agent completes a traversal of link ij, it updates the estimate of time cost C_{ij}^t as follows:

$$C_{ij}^t = (1-\lambda_t)C_{ij}^{t-1} + \lambda_t T_t, \tag{3.9}$$

where T_t is the amount of time that the agent just took to traverse link ij, and $\lambda_t \in [0,1]$ is a learning rate which is given as follows:

$$\lambda_t = \max\left(\frac{\lambda}{1+\frac{\kappa_{ij}^t}{20}}, 0.1\right), \tag{3.10}$$

where λ is a random number in the range $[0.2, 0.6]$ and κ_{ij}^t is the number of times that the agent has traversed link ij up to time t.

Selecting a link to traverse

After arriving at node i, and selecting its goal node $g \in N$, the agent must select which link to traverse. This is done differently for each agent. Let

$$Q_{ik}^t = \max_{k \in N_i}\left(C_{ik}^t + T_{ik}^t + G_{kg}^t\right). \tag{3.11}$$

The learning agent uses Wierings' Max-Boltzmann Exploration (or Softmax exploration) [50], wherein the agent selects to go to node $j \in N_i$ with the following probability:

$$Pr(j) = \frac{e^{Q_{ij}^t/\tau_t}}{\sum_{k \in N_i} e^{Q_{ik}/\tau_t}}, \tag{3.12}$$

where τ_t is the temperature parameter given by:

$$\tau_t = \frac{1.0}{5 + \frac{1}{2}\sum_{k \in N_i} \kappa_{ik}^t}. \tag{3.13}$$

Note that the agent selects each adjacent link with nearly equiprobable chances initially because of the high temperature value. The temperature parameter decreases over time with κ_{ik}^t increasing. As temperature decreases and as the distance between the Q-values increases, the probability that the agent selects greedily increases.

3.3.2 The Static Model

A static agent behaves in a similar manner as the learning agents except that the static agent does not update its estimate of time cost based on its experiences and it uses a different method to select a link. Static agents only respond to toll changes on each link. They do not learn the traffic status of each link.

Selecting a goal

A static agent uses the same method as the learning agents to select a goal node.

Estimating time to traverse a link

The estimated time cost for a static agent to traverse link ij is determined as follows:

$$C_{ij}^t = \frac{d_{ij}}{V_{ij}}. \tag{3.14}$$

Selecting a link to traverse

A static agent selects the node $j \in N_i$ (or link ij) to visit next as follows:

$$j = \arg\max_{k \in N_i} Q_{ik}^t. \tag{3.15}$$

where Q_{ik}^t is defined in Equation 3.11.

CHAPTER 4

Regulating MAS with Unlimited Interventions

We performed a user study[1] to investigate whether people are able to create effective interventions in Jiao Tong. We start by proposing four preliminary hypotheses, and then present the experimental design, followed by a discussion of the results.

4.1 Hypotheses

In this section, we propose four hypotheses with respect to Jiao Tong.

4.1.1 Hypothesis 1

In Jiao Tong, static agents do not adapt to the behaviors of other agents. They estimate the time costs to traverse a link solely based on the length of the link and the maximum velocity allowed on that link. These are fixed values that do not change during the game. Without human interventions, the tolls for entering each

[1]This study was approved by the Masdar Institute Human Subject Research Ethics Committee and was authorized as study HSREC-2013-4.

link are also fixed. If a node offers more incentives than the others, then there will be more agents choosing to traverse links that lead to it. If the number of agents running on the link(s) exceeds the capacity of the link(s), then the traffic will be extremely slow and the running throughput of the network will be decreased.

Learning agents, on the other hand, learn estimates of the traversal time of each link based on experiences. They adapt to the changing environment (e.g., both tolls and the behaviors of others). Though they do not have a global view of the environment, we consider learning agents to be less myopic and more capable of complex decision-making than static agents because these learning agents obtain better abstractions of the environment (the traffic situation of the network). Without human interventions, we believed that learning agents will outperform static agents in terms of running throughput. Hence, we made the following hypothesis:

Hypothesis 4.1 *In Jiao Tong, societies consisting of learning agents will perform better than societies consisting of static agents when no regulations are supplied.*

4.1.2 Hypothesis 2

We believed that it would not improve the performance of static agents significantly by human interventions. Because it seemed to us that it would not be difficult for regulators to learn and discern the preferences of static agents through observation and interaction with the agents. People could then figure out how to set tolls on each link so that traffic congestion or jams could be avoided and the running throughput could be increased substantially. Thus, we had the following hypothesis:

Hypothesis 4.2 *In Jiao Tong, human interventions will improve the performance of societies of static agents.*

4.1.3 Hypothesis 3

Though we anticipated that, in the absence of interventions, learning agents would outperform static agents in Jiao Tong, we did not believe that societies of learning agents would perform optimally without interventions. When learning agents make decisions (e.g., selecting a goal node, choosing a link to traverse), they do not merely respond to tolls on each link or the incentives they will receive after arriving at a node. They also adapt to the continuously changing traffic situations on each link. Consider a learning agent m at node B in Figure 4.1. It knows that the toll on link $\vec{BC}$ is higher than that on link $\vec{BD}$. But if agent m has learned that it will be very likely to take a much longer time to pass through $\vec{BD}$, it may select $\vec{BC}$ rather than $\vec{BD}$. In this case, regulators' interventions may not be effective. This is because it is rather difficult for regulators to establish an appropriate model of the learning agents in a short time since these agents' behaviors keep changing. Without good understanding of the agents' collective behaviors, regulators may not be able to create effective interventions. If inappropriate interventions are given, the performance of learning agents will be diminished. Therefore, we anticipated the following hypothesis:

Hypothesis 4.3 *In Jiao Tong, human interventions will decrease the performance of societies of learning agents.*

4.1.4 Hypothesis 4

From Hypothesis 4.2 and Hypothesis 4.3, we also hypothesized the following:

Hypothesis 4.4 *In Jiao Tong, humans will be better at regulating static agents than learning agents.*

An overview of the hypothesized relative running throughput for the four groups: learning model with interventions, static model with interventions, learning model

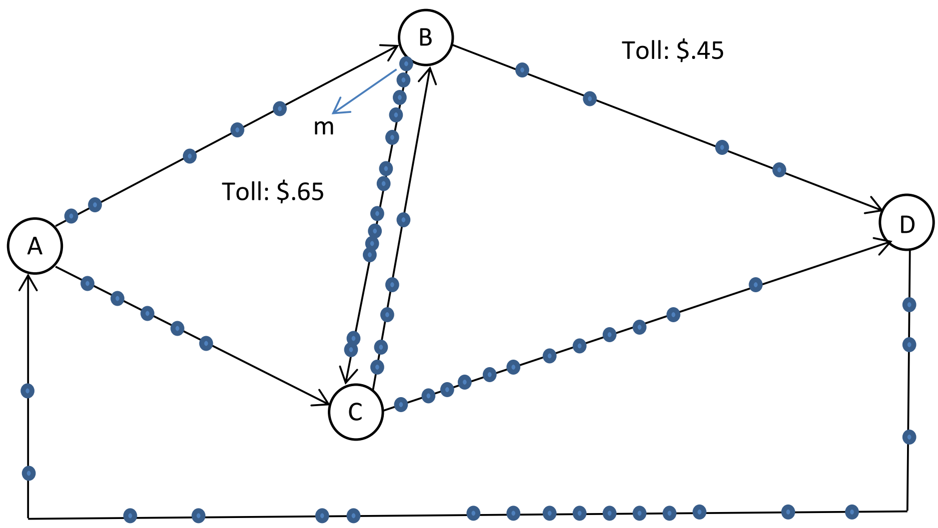

Figure 4.1: A scenario of the game when human intervention is not effective. A learning agent m selects link $\vec{BC}$ even though the toll on $\vec{BD}$ is less than the toll on $\vec{BC}$.

without interventions and static model without interventions is shown in Figure 4.2. These values are based on the four hypotheses discussed previously.

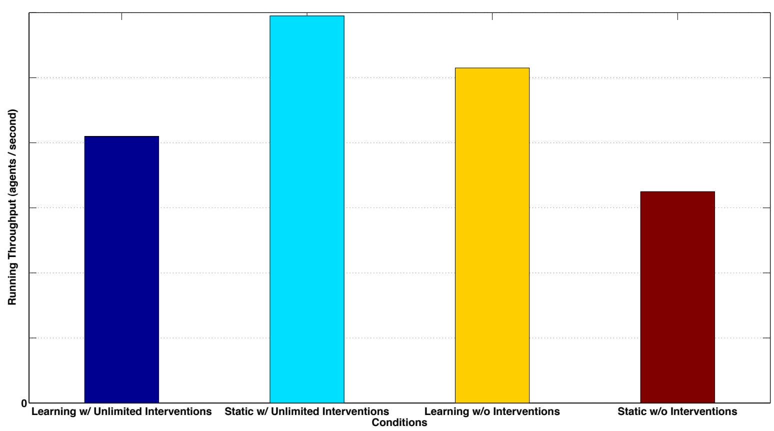

Figure 4.2: Relative hypothesized performance under 4 different scenarios: learning model with interventions, static model with interventions, learning model without interventions and static model without interventions.

4.2 Experimental Design

To test the hypotheses, we designed a *graphical user interface* (GUI) which simulated a transportation system and enabled users to create real-time interventions by changing tolls on each link. A snapshot of the GUI is shown in Figure 4.3. In this game, human subjects acted as regulators who followed the rules and constraints described in Section 3.2. They were able to change the tolls without limits (except the constraint on the scope of the tolls) if necessary.

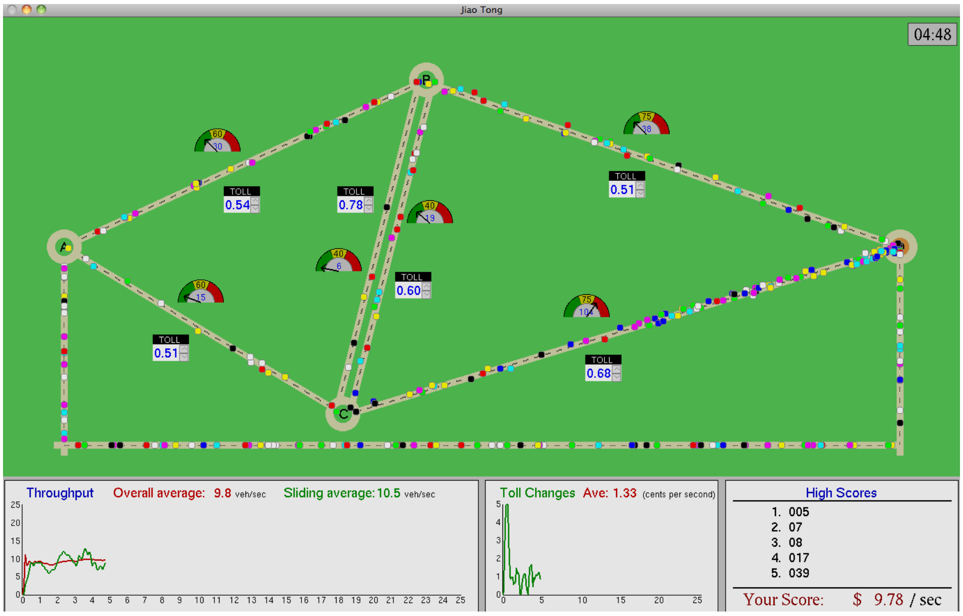

Figure 4.3: The GUI used in the experiment on regulating MAS using unlimited interventions.

4.2.1 Subjects

In this study, we performed two control experiments: *learning model without interventions* (LM) and *static model without interventions* (SM), where no human subjects were involved in. We also conducted two treatment experiments: *learning model with unlimited interventions* (LMU) and *static model with unlimited interventions* (SMU). 22 subjects (with normal visual perception) were recruited from

Masdar Institute to participate in an experiment at the Masdar Institute's Interactive Artificial Intelligence Laboratory. Among them, 11 subjects gave regulations for the learning models and the other 11 subjects were tested with the static models. The subjects were assigned to either the learning model (LMU) or the static model (SMU) with equal probability, and were not aware of the decision models used by the agents.

Among the participants in LMU group, ten subjects were male and one subject was female; 18.2% of the subjects (two subjects) held doctorate degrees, 27.3% of them (three subjects) had master's degrees, the rest of them (six subjects) owned bachelor's degrees. In the SMU group, there was no female participant. 27.3% of the subjects (three subjects) were masters and the remaining (eight subjects) were bachelors. All the subjects in both LMU and SMU were between the ages of 21 to 40 and had science or engineering backgrounds.

4.2.2 Experiment Protocol

The time constraint for a Jiao Tong game was 25 minutes. To investigate whether people were able to learn and figure out agents' behavioral models (learning or static) and preferences[2], node *B* was preferred by more agents than the other three nodes (A, C and D). All the other three nodes offered equal incentives to agents who reached them. This rule also applied to the control groups where no interventions were given.

In the control experiments, the tolls on each link were set to $0.50. No human participation or interventions were allowed in these conditions.

In the treatments, human interventions were needed. For each subject, the following procedure was rigorously followed:

1. The subject was asked to sign a consent form which described the rights of

[2]Here, preferences refer to the preferred nodes.

the subjects, the purpose of the study, the procedures of the experiments and the confidentiality of the data collection from the subjects.

2. The subject completed a demographic survey.

3. The subject was introduced to the system used in the experiment by the study administrator. Then the subject was trained on the system until he/she felt comfortable using the system. Note that the system used in this training session was different from the real experiments in order to avoid biases. In this system, agents did not react to the tolls nor adapt to the traffic situations on each link. They randomly selected their goal nodes and the links for traversing. The subject was also informed that this system was for training only. After training, the subject was tested to ensure that he/she had understood the task in the experiment and was familiar with the operations.

4. The subject completed a post-training survey. This survey was designed to gather information about the subjects' opinions with respect to interventions and agent autonomy.

5. The subject participated in a 25-minute Jiao Tong game with an assigned model (learning or static).

6. The subject completed a post-experiment survey, including questions about the subject's experience with the system.

During the test, the running throughput, the sliding throughput, the toll changes that the subjects made, and the status of each link were logged.

4.3 Results

In this section, we present an analysis of the results obtained from the user study. The main objectives were to evaluate whether the hypotheses proposed in Section 4.1 held or not, and to investigate the performance of the real-time interventions created by subjects in Jiao Tong. We conducted a series of one-way ANOVA (short for *Analysis of Variance*) tests on the mean values of each metric (running throughput, sliding throughput or toll changes) for the four groups (LMU, SMU, LM and SM) to find whether there existed statistically significant differences ($\alpha = 0.05$) between them.

4.3.1 Testing Hypothesis 1

We ran both the learning model and the static model for 11 times without human controls. Figure 4.4 shows a comparison of the overall running throughput for the two groups without intervention (LM and SM). The means of the running throughput (agents / second) for samples in LM and SM are 10.81 and 6.12, respectively. A one-way ANOVA demonstrates that the difference between the running throughput of LM and that of SM is highly significant ($F(1,20) = 60.73$, $p < .001$). Learning agents performed significantly better than static agents without human interventions in Jiao Tong. Therefore, Hypothesis 4.1 was verified to be true.

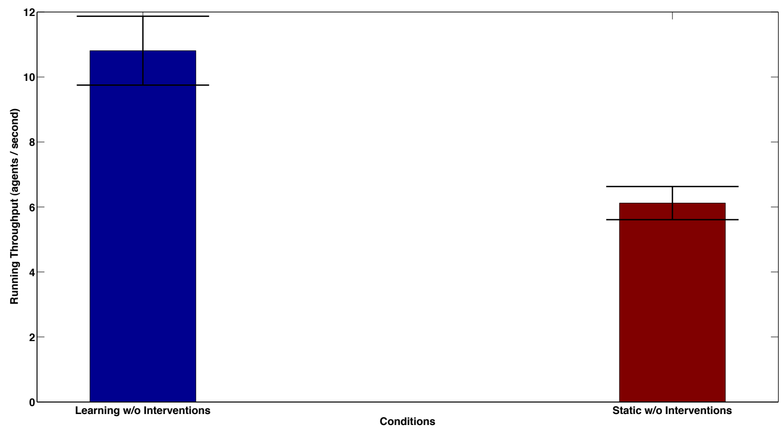

Figure 4.4: A comparison of running throughput between group LM (learning model without interventions) and group SM (static model without interventions). (Error bars show a 95% confidence interval on the mean.)

4.3.2 Testing Hypothesis 2

We next evaluate whether human interventions improve the performance of the static agents. To do this, we compared the performance of SMU with the performance of SM. As seen in Figure 4.5, the average running throughput in SMU (mean = 9.72, margin error = 1.37) is obviously higher than that in SM (mean = 6.12, margin error = 1.06). A one-way ANOVA shows a statistical difference between the running throughput of SMU and SM (F(1,20) = 23.44, $p < .001$). This implies that human interventions improved the performance of static agents significantly in Jiao Tong. Thus, Hypothesis 4.2 was accepted based on the experimental results.

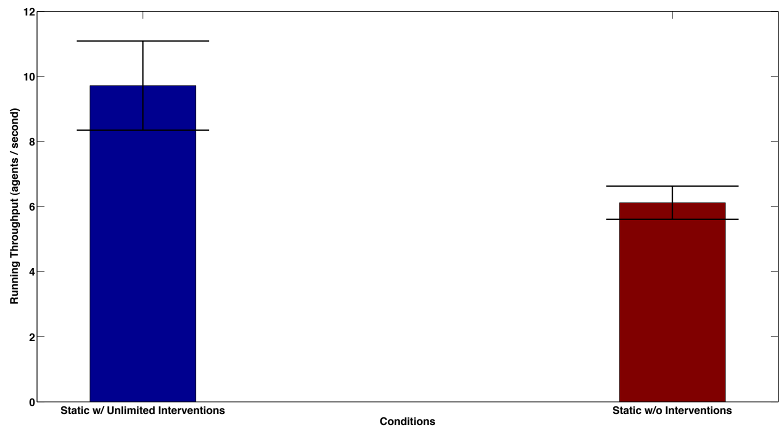

Figure 4.5: A comparison of running throughput between static model with interventions (SMU) and static model without interventions (SM). (Error bars show a 95% confidence interval on the mean.)

However, post-experiment surveys show that just three out of the 11 subjects in SMU realized that node *B* was preferred by more agents than the other nodes. Only two subjects believed that the agents in the network were not adapting. But none of the subjects were able to recognize both facts. 63.4% subjects (seven subjects) thought the system would converge to a congestion-free state through interventions. Among the seven subjects, six people managed to make the system congestion-free for more than three minutes uninterruptedly during the game, but only three successfully maintained that until the end of the game. One subject kept the system congestion-free from $t = 467s$ until the end ($t = 1500s$). He scored the highest among the 22 subjects (14.93 agents / second) though he mistook the static agents as learning agents. But, he understood agents' preferences for node B and adjusted the tolls properly.

Human interventions helped improve the performance of societies of static agents significantly. However, the improvement of performance did not live up to our expectations. We observed that many subjects frequently changed the tolls during game instead of attempting to wait and observe the agents' behaviors.

4.3.3 Testing Hypothesis 3

We next evaluate the impact of interventions on learning agents. A comparison of the running throughput between LMU and LM is illustrated in Figure 4.6. Experimental result shows that the average running throughput in LMU (mean = 10.73, $p = .89$) is slightly lower than that in LM (mean = 10.81, $p = 1.37$). A one-way ANOVA shows no statistically significant difference between them (F(1,20) = .01, $p = .911$). This indicates that human interventions did not appear to decrease the performance of learning agents in Jiao Tong. However, it did not appear to improve their performance either. So Hypothesis 4.3 was not confirmed with this experiment.

Post-experiment surveys indicate that only two subjects recognized that node B was preferred by more agents than the other nodes. 54.5% subjects (six subjects) noticed that the agents in the experiment were adapting. However, only one subject was able to point out both phenomena correctly. 81.8% subjects (nine subjects) believed the system would converge to some congestion-free states over time. Eight

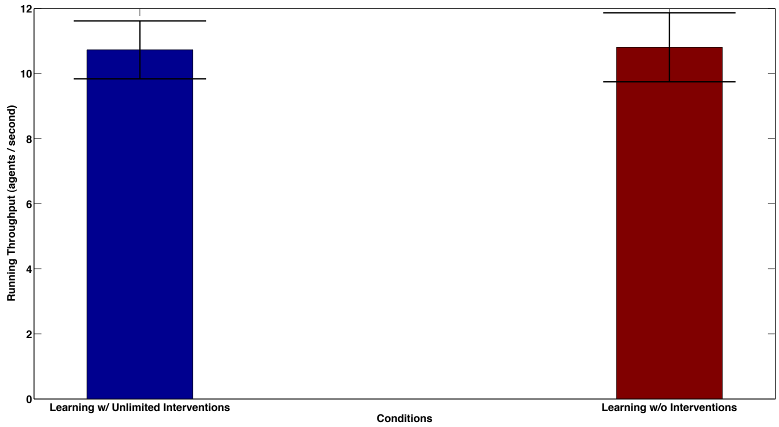

Figure 4.6: A comparison of running throughput between learning model with interventions (LMU) and learning model without interventions (LM). (Error bars show a 95% confidence interval on the mean.)

of them successfully made the system congestion-free for more than three minutes by changing tolls. Among the eight participants, three subjects failed to maintain the dynamic equilibrium until the end of the game. Through our observation, all these three subjects and another two subjects who failed to reach a congestion-free state frequently adjusted the tolls on the links to regulate the agents during the game.

4.3.4 Testing Hypothesis 4

The previous results indicate that on average people could create effective interventions to improve the performance of societies consisting of static agents, but they failed to design regulations to improve the performance of societies of learning agents. This infers that people were better at regulating static agents than learning agents in Jiao Tong. Thus, Hypothesis 4.4 held based on the experimental results. Figure 4.7 shows a comparison of the average running throughput between the control groups (LM and SM) and the experimental groups (LMU and SMU). However, the performance of static agents with interventions was not higher than the performance of learning agents with interventions.

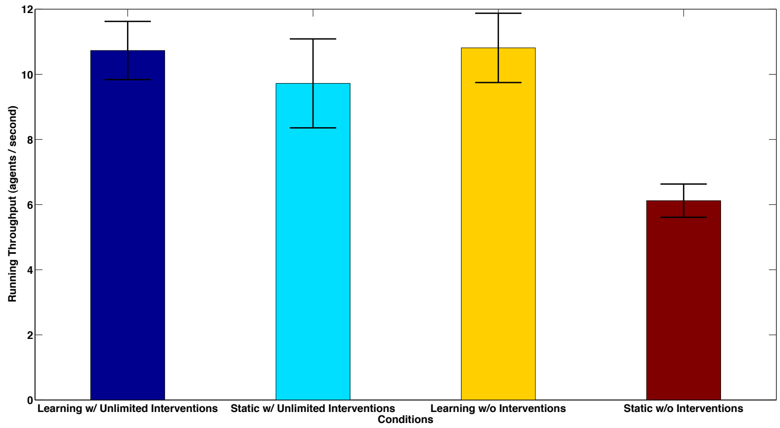

Figure 4.7: A comparison of the average running throughput between the control groups (LM and SM) and the experimental groups (LMU and SMU) (t=1500s). (Error bars show a 95% confidence interval on the mean.)

We compared the average running throughput over the duration (1500 seconds) of the game for all the four groups, of which two are control groups and the other two are treatments. From Figure 4.8, we can see that the running throughput went down shortly after the game started. The reasons are described as follows: first, all types of agents behaved greedily and had little or no experience of the network at the very beginning of the game, so it was not surprising to have traffic jams on some link(s); second, it took time for the congested road(s) to be cleared out, even if some subjects started adjusting the tolls on each link as soon as the game began. After reaching the lowest points, the average running throughput for all groups except SM increased gradually. For the static model without intervention, the average running throughput kept steady at approximately 6.1 agents per second.

From Figure 4.8, we observe that the average running throughput of LMU grew faster than that of LM and SMU from $t = 1200s$ to $t = 1500s$. This appeared to be due to increased throughput in the last five minutes of the game. Thus, we

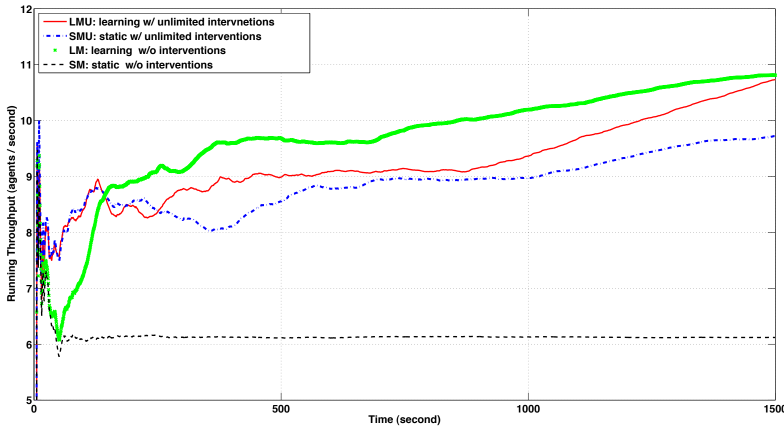

Figure 4.8: A comparison of average running throughput over time between the control groups (LM and SM) and the experimental groups (LMU and SMU).

might expect significant differences in the average sliding throughput from 1200s to 1500s between the three groups. Moreover, the average sliding throughput of SMU is obviously less than the other three groups according to Figure 4.9. We run a one-way ANOVA to determine whether these judgments are correct or not. As expected, there exist statistically significant differences among the four groups (F(3,40) = 12.83, $p < .001$). Through pairwise comparisons, the average sliding throughput of SMU in the last five minutes is significantly different from that of LMU (F(1, 20) = 4.68, $p = .043$). Since learning agents adapt to the environment, the accuracy of their estimation on the link costs increased over time. There is also a highly significant difference between SMU and SM (F(1,20) = 33.74, $p <$ .001). However, no statistically significant difference is detected between LMU and LM (F(1,20) = 1.06, $p = .316$). Figure 4.10 demonstrates the mean values of the average sliding throughput from 1200s to 1500s for the four groups with a confidence of 95% on the mean.

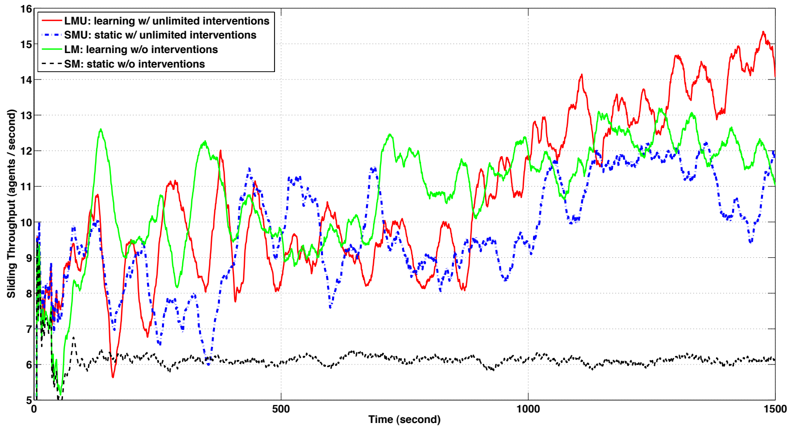

Figure 4.9: A comparison of sliding throughput over time between the control groups (LM and SM) and the experimental groups (LMU and SMU).

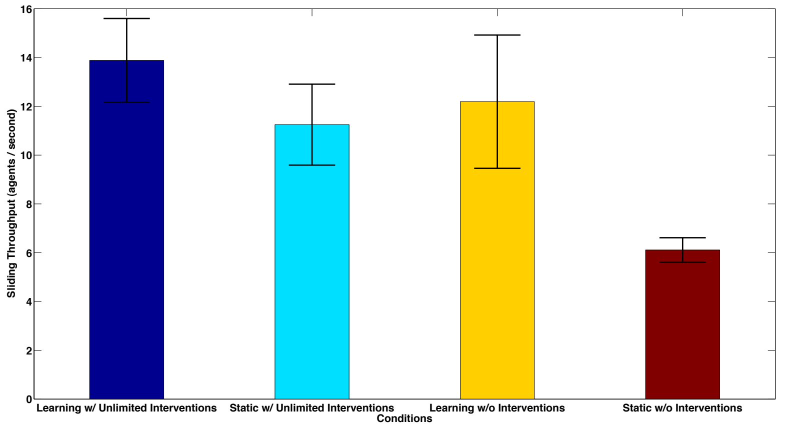

Figure 4.10: A comparison of average sliding throughput from 1200s to 1500s between the control groups(LM and SM) and the experimental groups (LMU and SMU). (Error bars show a 95% confidence interval on the mean.)

4.3.5 Discussion

While there was a difference in the different treatments, most subjects were unable to create effective regulations within 25 minutes in both scenarios. We are interested in determining what can be done to help users create better interventions.

Figure 4.11 shows the average amount of the total toll changes made in a Jiao Tong game (t = 1500s) for LMU and SMU. They are \$37.95 and \$45.62, respectively. Nevertheless, there is no statistically significant differences on the toll change between the two groups (F(1,20) = .54, $p = .47$). In other words, subjects used the same degree of interventions (similar amounts of toll changes) for regulating learning agents as they did for managing static agents in Jiao Tong.

During the user study, we observed that many of the participants (particularly those who performed poorly) changed tolls very often without waiting to observe how the agents reacted to toll changes. We believed that this could have negatively affected system throughput.

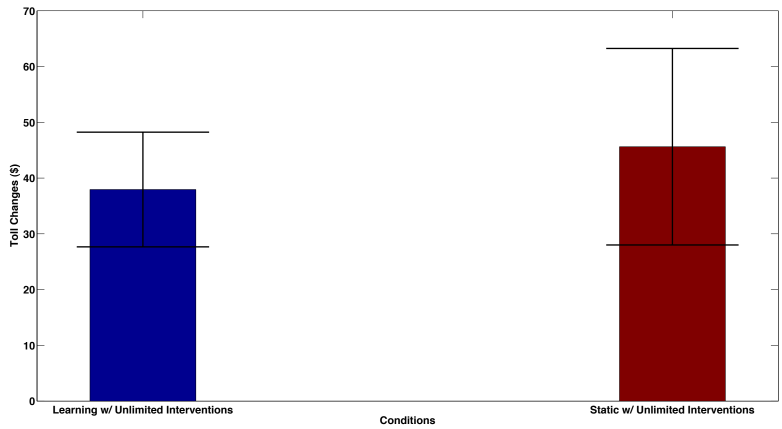

Figure 4.11: A comparison of toll changes between LMU and SMU. (Error bars show a 95% confidence interval on the mean.)

Figure 4.12 illustrates the toll change rates over time in the game for LMU and SMU. As seen from this figure, the toll change rates of both groups evolved in similar pattern until $t = 1200s$. From 1200s to 1500s, it seemed that subjects in SMU tended to change tolls more frequently than the participants in LMU. Nevertheless, the one-way ANOVA test shows no significant difference between them (F(1,20) = 2.29, $p = .146$). This indicates that people used similar amounts of toll changes in the last five minutes of the game for both models.

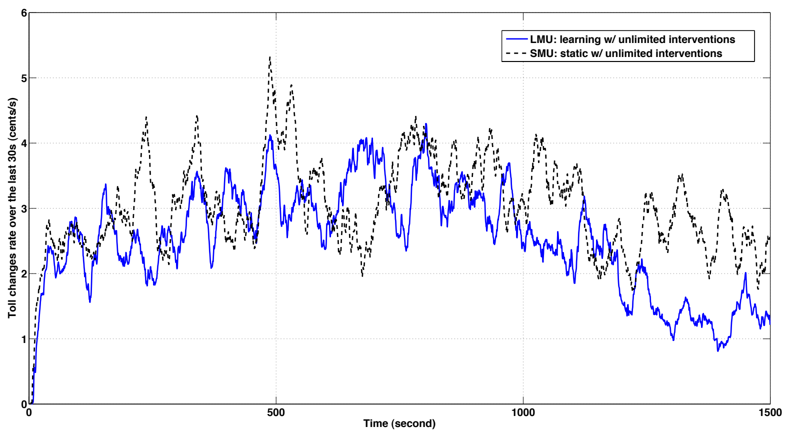

Figure 4.12: A comparison of toll change rates over time between LMU and SMU.

Figure 4.13 shows the correlation between the toll changes and the running throughput in LMU and SMU. A Pearson's r correlation test (Pearson product-moment correlation coefficient [32])[3] indicates a very strong negative relationship between the toll changes and the running throughput in SMU ($r(9) = -0.806, p < .01$). However, there is no statistically significant relationship between the two variables in LMU ($r(9) = -0.533, p > .05$). While this is only a correlation, we hypothesize that subjects were unable to create effective interventions because they chose to keep changing tolls rather than observing the agents' behaviors. The subjects would have performed better if they limited how often they changed tolls. In Chapter 5, we designed another two groups of experiments to test this hypothesis.

[3] We used the following interpretation of the size of a correlation based on the guidelines in [2]: None (-0.09 to 0.0), Weak(-0.3 to -0.1), Moderate (-0.5 to -0.3), Strong (-0.7 to -0.5), Very Strong (-1.0 to -0.7).

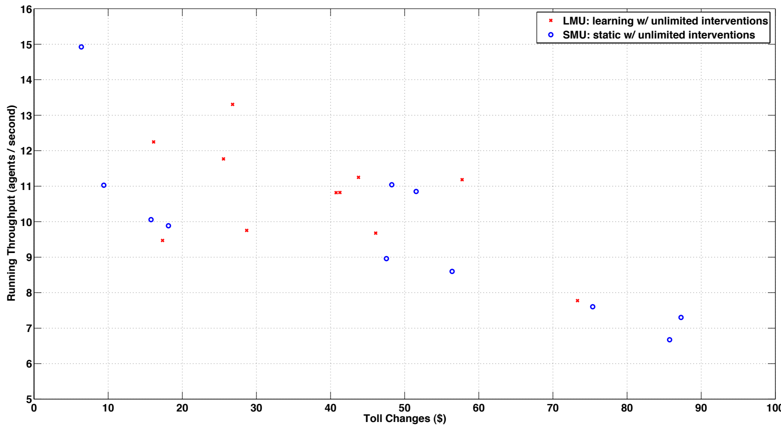

Figure 4.13: The correlation between the toll changes and the running throughput in LMU and SMU.

CHAPTER 5

Regulating MAS with Limited Interventions

In this chapter, we present a case study on regulating MAS by using constrained interventions. In this case study, the amount of toll changes the subjects were allowed to make was limited. We first introduce two hypotheses based on the discussion in Chapter 4. We then describe the experimental protocol in this user study. Finally, we present the experimental results.

5.1 Hypotheses

In this section, we propose two hypotheses on limited interventions in Jiao Tong.

5.1.1 Hypothesis 5

In the user study discussed in the previous chapter, we observed that people who performed poorly in regulating agents in Jiao Tong changed the tolls frequently. We anticipated that if we limited how often they could change the tolls, they would have been forced to observe the preferences of the agents more carefully. Since static agents do not adapt, it would not take too much efforts for them to figure out agents' preferred node(s) correctly. Once they identified the preferences of agents, they could adjust the tolls more wisely and more efficiently. Therefore, a significant increase in the performance of static agents was expected, when users were limited in the amount of toll changes they were allowed to make. To summarize, we had the following hypothesis:

Hypothesis 5.1 *In Jiao Tong, the performance of societies of static agents will be improved significantly by limiting the toll changes regulators can make, compared to the performance of societies of static agents when regulators can make unlimited toll changes.*

Figure 5.1 represents the anticipated relative performance between static model with unlimited interventions and static model with limited interventions as stated in Hypothesis 5.1.

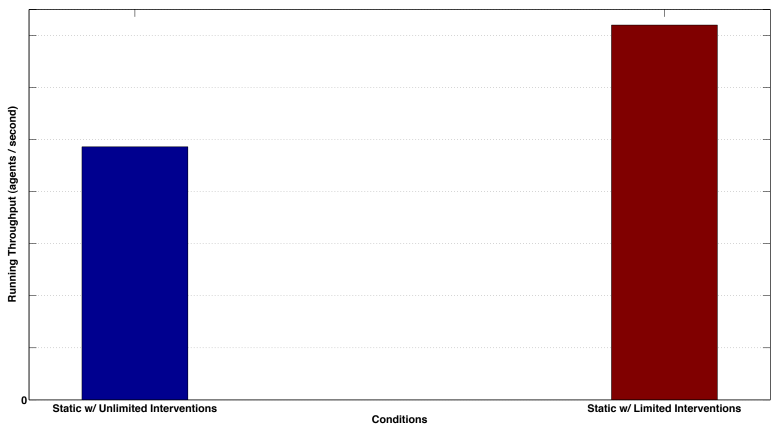

Figure 5.1: Anticipated relative performance between static model with unlimited interventions and static model with limited interventions based on Hypothesis 5.1.

5.1.2 Hypothesis 6

We also foresaw a slight increase in the performance of learning agents when toll changes were limited. It seemed to us that regulators would perform better when they they had a better understanding of the dynamics (e.g., the preferences of agents and the traversing agents on each link) in Jiao Tong. However, it would be more difficult for regulators to figure out the correct preferences of learning agents than that of static agents. This was because the traffic situations on each link kept changing during the game when learning agents adapted to the dynamic environment. Once the toll on one link had been changed, the traffic situations in the network would also be changed. That is, the environment would be changed. Even with limited toll changes, it would still be difficult for regulators to work out the preferred node(s) correctly. Thus, we only expected a slight increase in the performance of learning agents as described in Hypothesis 5.2 and illustrated in Figure 5.2.

Hypothesis 5.2 *In Jiao Tong, the performance societies of learning agents will be improved slightly by limiting the usage of interventions, compared to the performance of societies of learning agents which are given unlimited interventions.*

5.2 Experimental Design

To investigate whether the performance of static agents and learning agents would be improved by limiting the usage of interventions, we performed two groups of human subject experiments: *static model with limited interventions* (SML) and *learning model with limited interventions* (LML). A snapshot of the GUI used in the experiments is shown in Figure 5.3.

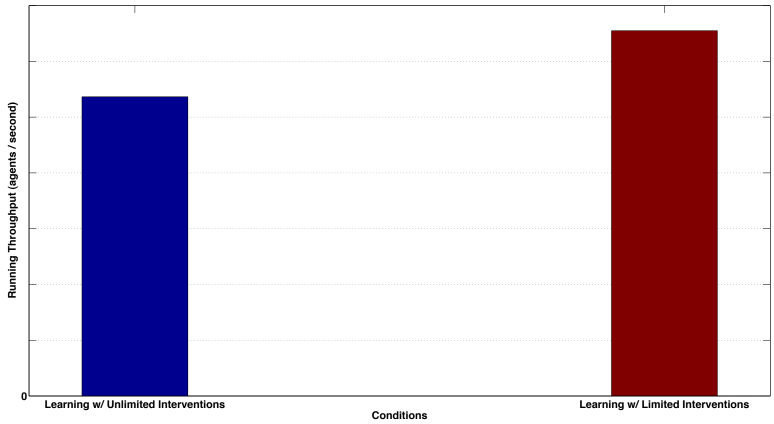

Figure 5.2: Anticipated relative performance between learning models with unlimited interventions and learning models with limited interventions based on Hypothesis 5.2.

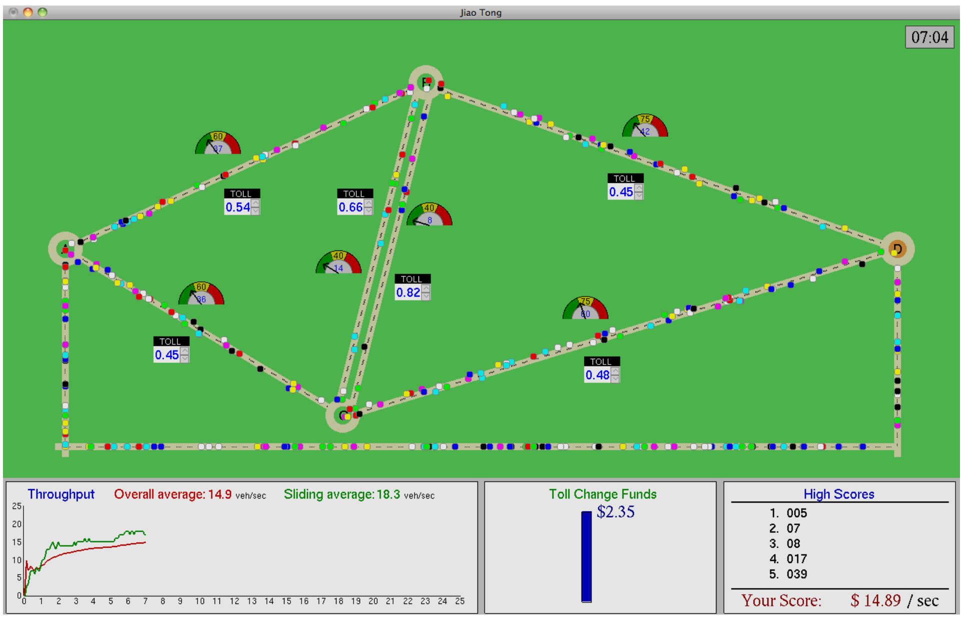

Figure 5.3: The GUI used in the experiment on regulating MAS with limited interventions.

5.2.1 Subjects

22 subjects ranging in age from 21-40 were recruited to participant in an experiment with either a static model or a learning model. All subjects were students or

staff at Masdar Institute.

Among the 11 participants placed in the SML group, eight of them were male and three subjects were female. One subject had a doctoral degree and four of them had master's degrees. The rest of them (six subjects) held bachelor's degrees. In the LML group, the ratio of males to females was also 8 : 3. Three of them had bachelor's degrees and one subject obtained a PhD degree. The other seven participants had master's degrees. All the subjects in both groups had science or engineering backgrounds.

5.2.2 Experimental Protocol

The experimental protocol was the same as described in Section 4.2.2 except that participants were limited in the amount of toll changes they were permitted to make.

At the beginning of the game, each subject had \$0.30 in his/her toll change fund. Once a subject increased or decreased the toll on a link by some amount (e.g., \$0.05), the toll change fund would be reduced by that amount (e.g., \$0.05). The toll change fund grew by \$0.007 per second during the game. In the game, the maximum amount of toll changes that a subject was allowed to make was \$10.8, which was substantially less than the average amount of toll changes made in the LMU (\$37.95) and SMU (\$45.62) groups. Even with much less amount of toll changes than the maximum amount allowed (\$10.8), it was also possible for subjects to make effective interventions. For instance, the subject who scored the highest among the 22 subjects in the user study described in the previous chapter only made \$6.38 toll changes.

If the subject used up the toll change fund, he/she was not allowed to change the tolls on the links. In such cases, the subject needed to wait for a few seconds until he/she had sufficient toll change fund.

5.3 Results

In this section, we perform statistical analyses of the experimental data collected from the user study to determine whether the hypotheses proposed in Section 5.1 are correct or not.

5.3.1 Testing Hypothesis 5

Figure 5.4 shows the average throughput obtained by the SML and SMU groups. The figure shows that the mean values of running throughput in group SML (mean = 10.80, margin error = .74) was slightly higher than that in group SMU (mean = 9.72, margin error = 1.37). However, a one-way ANOVA (F(1,20) = 1.84, $p =$ .19) failed to reject the null hypothesis[1], indicating that there is no statistically significant difference in the means of running throughput between the two groups of samples.

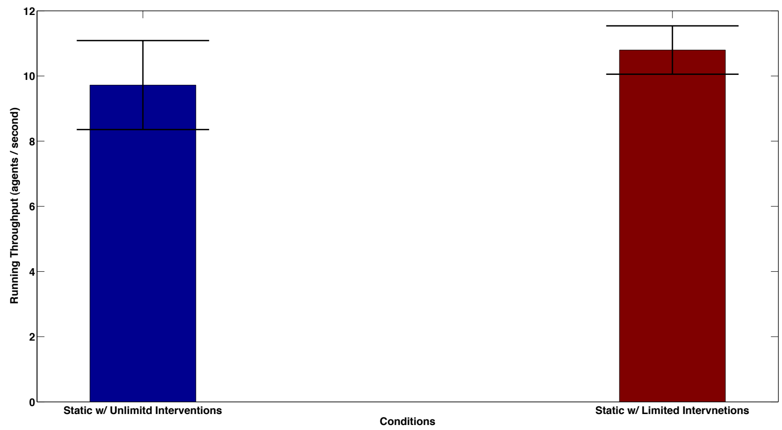

Figure 5.4: A comparison of the means of running throughput between SMU (static model with unlimited interventions) group and SML group (static model with limited interventions). (Error bars show a 95% confidence interval on the means.)

We also compared performance between the SML group and the SMU group

[1] $H_0 : \mu_1 = \mu_2$, where μ_i was the mean of the running throughput in group i, and $i \in \{SMU, SML\}$.

over time. As illustrated in Figure 5.5, the average running throughput of both groups dropped shortly after the game started and then increased gradually. This is because node B was preferred by more agents than other nodes, there was traffic congestion on link $\vec{BC}$ or link $\vec{CB}$ in the first few minutes of the game. Subjects in both groups needed to observe the dynamic of games in the very beginning. Thus, most of them made few toll changes. Even for the societies of agents where more toll changes were made, it still took time for the the traffic on the congested link(s) to be cleared out. So there was no substantial difference in the performance between the two groups in the first few minutes. Starting from $t = 300s$, the running throughput in the SML group (mean = 8.23, margin error = .99) was significantly higher than the throughput in the SMU group (mean = 10.73, margin error = 1.07). A one-way ANOVA shows a statistical significant difference in the throughput between the two groups when $t = 300s$ (F(1,20) = 11.35, $p = .003$). A comparison of the running throughput between the four groups at $t = 300s$ is shown in Figure 5.6.

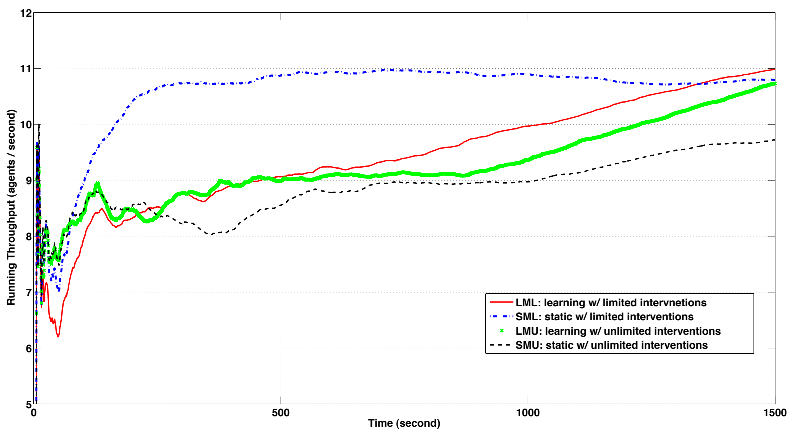

Figure 5.5: A comparison of the average running throughput over time between SMU, LMU, SML and LML.

After that, the running throughput in SMU increased slowly while the throughput in SML kept steady. But the performance in the SML group kept outperform-

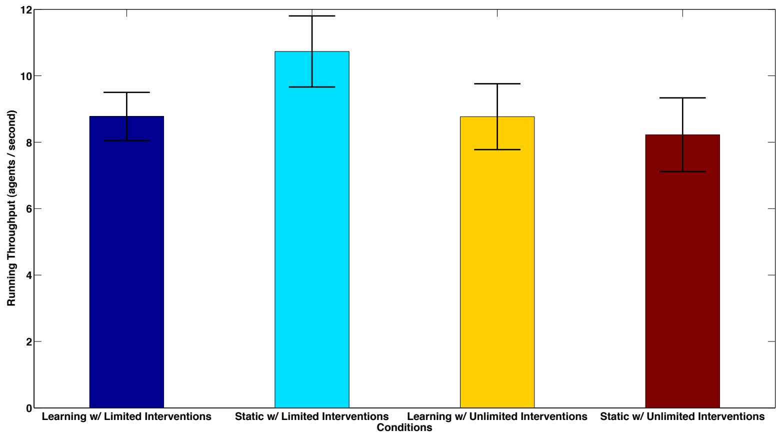

Figure 5.6: A comparison of the running throughput between SMU, LMU, SML and LML at $t = 300s$. (Error bars show a 95% confidence interval on the means.)

ing the performance in the SMU group until $t = 1100s$ (when $t = 1100s$, F(1,20) = 4.43, $p = .048$). Since then, the difference in the running throughput between the two groups became insignificant (e.g., when $t = 1200s$, F(1,20) = 2.91, $p = .104$; when $t = 1300s$, F(1, 20) = 2.2, $p = .154$.) until the end of the game. This may be due to the increase of sliding throughput from $t = 1100s$ as seen in Figure 5.7 in SMU.

The reason why the performance in SMU could slowly catch up with the performance in SML since $t = 300s$ is still unclear to us. However, it may be because subjects in the SML group were still trying to use up all the available toll change funds instead of observing the dynamics of the game carefully. In the last few minutes, subjects in both groups had accumulated some experience in regulating the static agents. Subjects in the SML group might have lacked toll change funds and could not create interventions when necessary. Subjects in the SMU group, however, had sufficient toll change funds and were able to change the tolls without constraints.

Post-experiment surveys show that five subjects noticed that node B was the

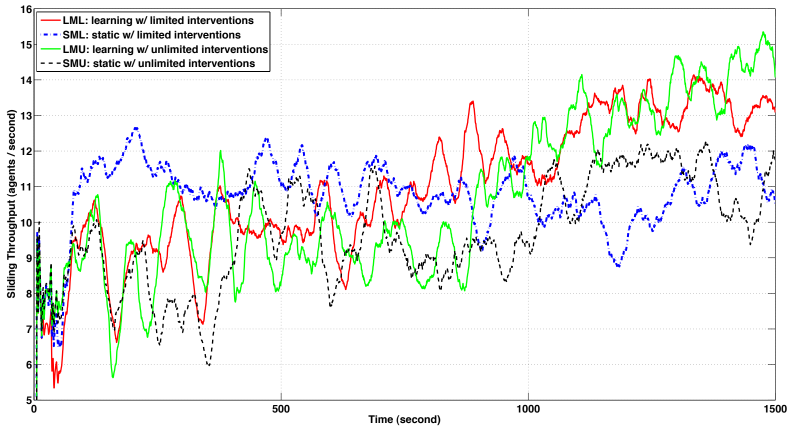

Figure 5.7: The comparison of average sliding throughput over time between SMU, LMU, SML and LML.

favored by more agents than the other nodes though all of them tried to observe the dynamics of the game (In the SML group, only three subjects identified the fact correctly). However, no subject was able to point out that the agents were not adapting. Four subjects maintained the congestion-free state until the end of the game (for more than three minutes uninterruptedly).

In summary, with the limits on the amount of toll changes, subjects were forced to observe the dynamics of the game for some time. However, it did not help all the subjects to identify the preferences of agents nor guarantee a better performance of the societies of the agents. It improved the performance of societies of static agents by restricting the amount of toll changes in the short-term (from 300s to 1100s in the game), but it failed to maintain the advantages in the long-term (e.g., 1500s). Therefore, Hypothesis 5.1 did not hold according to our experimental results.

5.3.2 Testing Hypothesis 6

We next investigate whether people could improve the performance of learning agents with limited interventions. We compared the running throughput of agents in LML with that in LMU. Experimental data (see Figure 5.8) shows that the average running throughput in LML (mean = 10.98 , margin error = .93) is very close to that in LMU (mean = 10.73 , margin error = .89). A one-way ANOVA (F(1,20) = .15, $p = .701$) shows no statistical difference exist between the means of running throughput in the two groups, indicating Hypothesis 5.2 was not true.

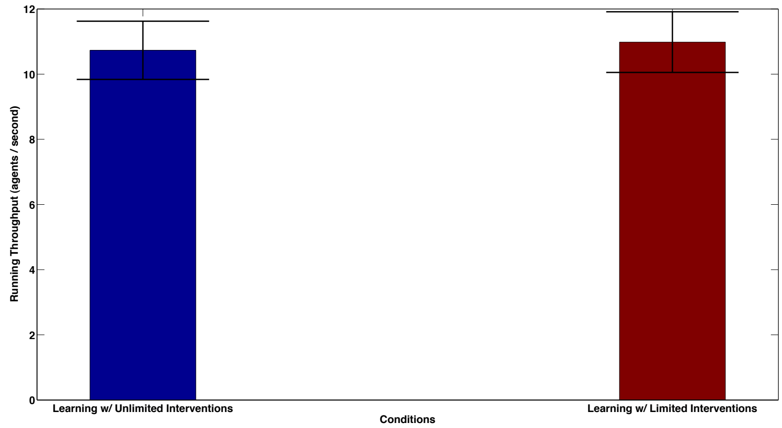

Figure 5.8: A comparison of means of running throughput between group LMU (learning model with unlimited interventions) and group LML (learning model with limited interventions). (Error bars show a 95% confidence interval on the means.)

The throughput in both the LML group and the LMU group dropped first and the increased slowly. The running throughput LML was quite near to that in LMU, though the performance in LML was slightly better than the latter group since $t = 600s$. Nevertheless, according to statistical tests, no significant difference in the throughput between the two groups existed during the game. This implied that it did not improve the performance of societies of learning agents by restricting the amount of toll changes made by the subjects.

Post-experiment surveys suggest that only three subjects were able to identify that node B was preferred by more agents that other nodes (In the LML group, two subjects identified the fact correctly). Eight participants thought the agents were adapting during the game. Five subjects made the systems congestion-free for more than three minutes uninterruptedly during the game and three of them maintained that until the termination of the game. Though people were forced to observe dynamics of the agents, they failed to figure out the preferences of learning agents, possibly due to the complexity of the problem. This indicates that it is not effective to attempt to increase the situation awareness [13] by solely limiting the usage of interventions in Jiao Tong. Subsequently, the performance of societies of learning agents will not be improved.

The average running throughput in LML even surpassed that in SML by the end of the game. From Figure 5.7, we can observe that the sliding throughput from $t = 1100s$ to $t = 1500s$ appeared to be apparently higher than the sliding throughput in SML. A one-way ANOVA shows that the mean throughput from $t = 1100s$ and $t = 1500s$ between the two groups is statistically different (F(1,20) = 4.32, $p = .05$).

We also compared the running throughput among SMU, LMU, SML and LML (as shown in Figure 5.9). It seems that there is no significant difference between them. A one-way ANOVA (F(3,40) = 1.22, $p = .317$) shows no statistically significant difference between them. However, from Figure 5.5, we can observe that the average running throughput in all groups dropped and then increased gradually shortly after the game started (no statistical difference detected among them). From 300s to 1100s, the performance in SML is substantially better than the performance in the other three groups. A one-way ANOVA shows that there is a statistical difference in the average throughput from 300s to 1100s among SMU, LMU, SML and LML (F(3,40) = 5.23, $p = .004$). Pairwise comparisons show that there exist statistical differences in running throughput between the following group pairs: SML and LML (F(1,20) = 10.54, $p = .004$), SML and LMU (F(1,20) = 10.39, $p = .004$), SML and SMU (F(1,20) = 9.43, $p = .006$). However, there is no statistical difference between SMU, LMU and LML (F(2,30) = .49, $p = .6197$).

From 1100*s* to the end of the game, no statistical difference exist among the four groups.

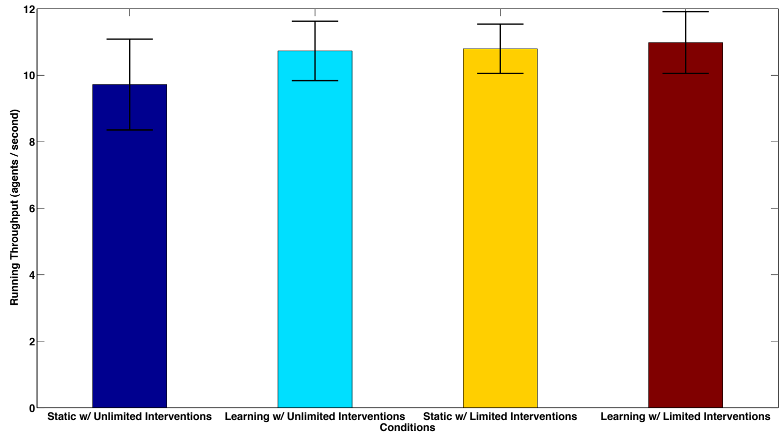

Figure 5.9: A comparison of the average running throughput between group SMU, LMU, SML and LML. (Error bars show a 95% confidence interval on the means.)

As seen in Figure 5.5, subjects in SML were able to achieve the highest running throughput early in the game. This might be because fewer subjects in SMU and LMU were attempting to observe the preferences of agents while more subjects in SML were forced to learn the preferences of the agents. Moreover, it might be easier for the subjects to figure out the preferences of static agents than that of learning agents after a few minutes' observation.

In conclusion, limiting interventions helped participants that regulated static agents in early stages of the game. However, these limitations failed to provide long-term benefits. Furthermore, limiting interventions did not help nor hurt the performance of participants that regulated learning agents.

CHAPTER 6

Conclusions and Future Work

Dynamic resource allocation (or dynamic resource control) in multi-agent systems has numerous real-world applications, especially in solving some very challenging problems related to sustainability such as traffic management problems in transportation systems, energy management problems in building systems and demand side management problems in power grids and water supply network.

These adaptive systems share several commonalities. First, resources in these systems are bounded or limited. Agents in these systems need to share the limited resources with their peers. For instance, a road can only accommodate a limited number of vehicles at the same time and drivers need to share the road with others. Second, agents in these systems are autonomous or self-motivated, which means they are not directly controlled by the regulatory entities, though they are subject to the regulations implemented by the regulators. For example, drivers in a transportation system may have to follow the traffic regulations. Otherwise, they may be subject to monetary penalties or legal sanctions. However, different drivers may have different preferred destinations or routes. Third, the preferences of agents are initially unknown to the regulators and are subject to change. In a transportation system, drivers may not be willing to reveal their true preferences of selecting destinations or routes to the transportation authorities due to privacy or other considerations. And they may adapt to the dynamic environment and change their preferences to maximize their utilities. Fourth, without proper intervention or management, these multi-agent systems may not meet societal goals. This is because agents only have partial or incomplete information of the environment or world, as they can only perceive their immediate surroundings. For instance, as the traffic demand approaches the capacity of a road, the speed of the traffic flow will slow down, which may result in some traffic congestion. Some drivers in the transportation network may not be able to know the traffic situation on that road. They may select this road for traversing. If the traffic on the road continues to increase, it is highly possible that the road will be jammed for some period of time, which may violate the interests of the whole in the transportation network. Therefore, it is essential and necessary to investigate principles of real-time interventions in these multi-agent systems.

In this work, we proposed an abstract transportation framework named *Jiao Tong* to model the dynamic resource allocation problems in the real world. We developed a user interface based on Jiao Tong, which enabled users to create real-time interventions to manage the transportation system.

We conducted two user studies to investigate whether people could effectively design interventions to manage resource-limited multi-agent systems in real-time. We bring the following conclusions according to our experimental results. First, societies of adapting agents perform substantially better than societies of non-adapting agents without human interventions, because adapting agents are considered to be less myopic and more capable of operating in complex and dynamic environments than non-adapting agents. Second, human interventions significantly improve the performance of societies of non-adapting agents. Since non-adapting agents do not adapt to the dynamic environment and they only respond to the interventions (toll changes in Jiao Tong), it is relatively easy for regulators to figure out the preferences of non-adapting agents through observation and interaction. However, when no limits on the usage of interventions are given, most people do not attempt to observe the dynamics of the problem (agents' behaviors in Jiao Tong). They tend to keep regulating the agents frequently. Third, human interventions do not help increase nor decrease the performance of societies of learning agents. Fourth, humans are better at regulating non-adapting agents than adapting agents. Fifth, with limits on the usage of interventions, people are forced to observe the dynamics of the problems. It improves the performance of societies of static agents in the short-term. However, it does not improve the situation awareness of all the people, nor help improve the performance in the long-term. Sixth, it does not help increase the performance of societies of learning agents by limiting the usage of interventions due to the complexity of the dynamics.

According to our experiments, we find that most people did not understand the dynamics well. They tended to keep changing tolls to make the congested links resume operating smoothly, while ignoring the preferences of agents, resulting in bad decision-making in interventions. So further investigation is left to address on effective methodologies to help regulators improve their understandings of agents' behaviors and help them create better interventions to regulate the MAS.

CHAPTER 7

Abbreviations

MAS Multi-agent Systems

BMS Building Management Systems

DRAP Dynamic Resource Allocation Problem

GUI Graphical User Interface

ANOVA Analysis of Variance

LM Learning Model without Interventions

SM Static Model without Interventions

LMU Learning Model with Unlimited Interventions

SMU Static Model with Unlimited Interventions

SML Static Model with Limited Interventions

LML Learning Model with Limited Interventions

Bibliography

[1] J.A. Adams and R. Paul. Human supervisory control of multiple mobile agents. In *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on*, volume 4, pages 3298–3303. IEEE, 1995.

[2] Per Ahlgren, Bo Jarneving, and Ronald Rousseau. Requirements for a cocitation similarity measure, with special reference to pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, 54(6):550–560, 2003.

[3] Alvin AuYoung, Brent Chun, Alex Snoeren, and Amin Vahdat. Resource allocation in federated distributed computing infrastructures. In *Proceedings of the 1st Workshop on Operating System and Architectural Support for the On-demand IT InfraStructure*, volume 9, 2004.

[4] Russell Bent and Pascal Van Hentenryck. The value of consensus in online stochastic scheduling. *ICAPS 2004*, 2004.

[5] S. Borenstein. *Time-varying retail electricity prices: Theory and practice*. University of Chicago Press, 2005.

[6] Gruia Calinescu, Amit Chakrabarti, Howard Karloff, and Yuval Rabani. Improved approximation algorithms for resource allocation. *Integer Programming and Combinatorial Optimization*, pages 401–414, 2006.

[7] Julia Chuzhoy, Rafail Ostrovsky, and Yuval Rabani. Approximation algorithms for the job interval selection problem and related scheduling problems. *Mathematics of Operations Research*, 31(4):730–738, 2006.

[8] Balázs Csanád Csáji. Adaptive resource control. 2008.

[9] Balázs Csanád Csáji and László Monostori. Adaptive stochastic resource control: a machine learning approach. *Journal of Artificial Intelligence Research*, 32(1):453–486, 2008.

[10] ML Cummings and PJ Mitchell. Operator scheduling strategies in supervisory control of multiple uavs. *Aerospace science and technology*, 11(4): 339–348, 2007.

[11] R.M. Dawes, D. Faust, and P.E. Meehl. Clinical versus actuarial judgment. *Science*, 243(4899):1668, 1989.

[12] Edsger W Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.

[13] Mica R Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

[14] Joan Feigenbaum, Christos Papadimitriou, Rahul Sami, and Scott Shenker.

A bgp-based mechanism for lowest-cost routing. *Distributed Computing*, 18 (1):61–72, 2005.

[15] Aram Galstyan, Karl Czajkowski, and Kristina Lerman. Resource allocation in the grid using reinforcement learning. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pages 1314–1315. IEEE Computer Society, 2004.

[16] Aram Galstyan, Karl Czajkowski, and Kristina Lerman. Resource allocation in the grid with learning agents. *Journal of Grid Computing*, 3(1):91–100, 2005.

[17] D. Gatherer. The spread of irrational behaviours by contagion: An agent micro-simulation. *Journal of Memetics-Evolutionary Models of Information Transmission*, 6, 2002.

[18] R.L. Goldstone and M.A. Janssen. Computational models of collective behavior. *Trends in Cognitive Sciences*, 9(9):424–430, 2005.

[19] Daniel Grosu and Anthony T Chronopoulos. Algorithmic mechanism design for load balancing in distributed systems. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 34(1):77–84, 2004.

[20] Dirk Helbing. Agent-based simulation of traffic jams, crowds, and supply networks. In *Proceedings of the IMA Hot Topics Workshop*, 2003.

[21] John Hershberger and Subhash Suri. Vickrey pricing in network routing: Fast payment computation. In *Proc. of the 42nd IEEE Symposium on Foundations of Computer Science*, pages 252–259, 2001.

[22] Birgit Heydenreich, Rudolf Müller, and Marc Uetz. Decentralization and mechanism design for online machine scheduling. *Algorithm Theory–SWAT 2006*, pages 136–147, 2006.

[23] Dorit S Hochbaum. *Approximation algorithms for NP-hard problems*. PWS Publishing Co., 1996.

[24] David S Johnson. Approximation algorithms for combinatorial problems. *Journal of computer and system sciences*, 9(3):256–278, 1974.

[25] Donald B Johnson. A note on dijkstra's shortest path algorithm. *Journal of the ACM (JACM)*, 20(3):385–388, 1973.

[26] P.M. Jones and C.M. Mitchell. Human-machine cooperative interaction in the supervisory control of a complex dynamic system. In *Systems, Man, and Cybernetics, 1991.'Decision Aiding for Complex Systems, Conference Proceedings., 1991 IEEE International Conference on*, pages 1301–1306. IEEE, 1991.

[27] D. Kahneman. Intuitive predictions: biases and corrective procedures. reprinted in kahneman, slovic, and tversky, judgement under uncertainty: Heuristics and biases, 1982.

[28] E. Kamar, S. Hacker, and E. Horvitz. Combining human and machine intelligence in large-scale crowdsourcing. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2012.

[29] Noa E Kfir-Dahav, Dov Monderer, and Moshe Tennenholtz. Mechanism design for resource bounded agents. In *MultiAgent Systems, 2000. Proceedings. Fourth International Conference on*, pages 309–315. IEEE, 2000.

[30] Harold Joseph Kushner and G George Yin. Stochastic approximation algorithms and applications. 1997.

[31] Ron Lavi and Chaitanya Swamy. Truthful mechanism design for multidimensional scheduling via cycle monotonicity. *Games and Economic Behavior*, 67 (1):99–124, 2009.

[32] Joseph Lee Rodgers and W Alan Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.

[33] Li Lin, Yu Zhang, and Jinpeng Huai. Sustaining incentive in grid resource allocation: A reinforcement learning approach. In *Cluster Computing and the Grid, 2007. CCGRID 2007. Seventh IEEE International Symposium on*, pages 145–154. IEEE, 2007.

[34] S.R. Musse and D. Thalmann. A model of human crowd behavior: Group inter-relationship and collision detection analysis. In *Computer Animation and Simulation*, volume 97, pages 39–51. Citeseer, 1997.

[35] Y. Nagar and T.W. Malone. Combining human and machine intelligence for making predictions. *MIT Center for Collective Intelligence Working Paper*, (2011-02), 2011.

[36] Noam Nisan and Amir Ronen. Algorithmic mechanism design. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 129–140. ACM, 1999.

[37] Ryan Porter. Mechanism design for online real-time scheduling. In *Proceedings of the 5th ACM conference on Electronic commerce*, pages 61–70. ACM, 2004.

[38] S. Ramchurn, P. Vytelingum, A. Rogers, and N. Jennings. Agent-based control for decentralised demand side management in the smart grid. In *Proceedings of the 10th International Conference on Autonomous Agents and*

Multiagent Systems, pages 5–12. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[39] H. Rose, A. Rogers, and E.H. Gerding. Mechanism design for aggregated demand prediction in the smart grid. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.

[40] H. Rose, A. Rogers, and E.H. Gerding. A scoring rule-based mechanism for aggregate demand prediction in the smart grid. 2012.

[41] Barna Saha. Approximation algorithms for resource allocation. 2011.

[42] T.B. Sheridan. *Telerobotics, automation, and human supervisory control*. The MIT press, 1992.

[43] S. Stein, E. Gerding, V. Robu, and N. Jennings. A model-based online mechanism with pre-commitment and its application to electric vehicle charging. 2012.

[44] D.J.T. Sumpter. The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1465):5–22, 2006.

[45] Gerald Tesauro et al. Online resource allocation using decompositional reinforcement learning. In *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*, volume 20, page 886. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.

[46] Sebastian Thrun. Lifelong learning: A case study. Technical report, DTIC Document, 1995.

[47] P. Trigo and H. Coelho. Collective-intelligence and decision-making. *Computational Intelligence for Engineering Systems*, pages 61–76, 2011.

[48] Hal R Varian. Economic mechanism design for computerized agents. In *First USENIX Workshop on Electronic Commerce*, pages 13–21, 1995.

[49] Vijay V Vazirani. *Approximation algorithms*. springer, 2004.

[50] Marco A Wiering. Explorations in efficient reinforcement learning. 1999.

[51] C.H. Yu, J. Werfel, and R. Nagpal. Collective decision-making in multi-agent systems by implicit leadership. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 3-Volume 3*, pages 1189–1196. International Foundation for Autonomous Agents and Multiagent Systems, 2010.