CONTROL OF CABLE NETWORKS

A Dissertation by

Haihong Gao

Master of Science, Wichita State University, 2002

Submitted to the Department of Electrical Engineering and Computer Science
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

December 2012

CONTROL OF CABLE NETWORKS

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy, with a major in Electric Engineering.

____

Ravi Pendse, Committee Chair

____

Edwin Sawan, Committee Member

____

John Watkins, Committee Member

____

Abu Asaduzzaman, Committee Member

____

Linda Kliment, Committee Member

Accepted for the College of Engineering

____

Zulma Toro-Ramos, Dean

Accepted for the Graduate School

____

Abu S.M. Masud, Interim Dean

DEDICATION

To Qingyu Xu, a professor at Xi'an Jiaotong University, who gave me confidence in my capability, which keeps me constantly pursuing my ambition

You have to have confidence in your ability, and then be tough enough to follow through.
— Rosalynn Carter

Success is not final, failure is not fatal: it is the courage to continue that counts.
— Winston S. Churchill

When someone tells me "no," it doesn't mean I can't do it, it simply means I can't do it with them.
— Karen E. Quinones Miller

ACKNOWLEDGEMENTS

I would like to thank my adviser, Dr. Ravi Pendse, for his guidance and constant support during my study for my PhD degree in electrical engineering.

I appreciate Dr. Patricia Best's help with the two published research papers regarding my dissertation.

I thank Dr. Edwin Sawan, Dr. John Watkins, Dr. Abu Asaduzzaman, and Dr. Linda Kliment for their participation in my PhD committee and for their precious time and valuable suggestions. Especially, I am grateful to Dr. Edwin Sawan for his suggesting me to work on cable networks.

I am grateful to the WSU interlibrary-loan staff for their help with my collecting research materials.

I thank my senior colleague, Mr. Amarnath Jasti, for his help during my working as a research assistant in Advanced Network Research Institute at WSU.

I thank WSU staff in department of electrical engineering and computer science, graduate school, university library, register's office, and etc. for their services.

ABSTRACT

A major issue for contemporary cable networks under flat pricing is that heavy users can cause severe network congestion and light users subsidize heavy users. This problem is exaggerated when network participants do not obey the protocols. From the perspective of control theory, cable networks are primarily open-loop systems that assume the participants are non-strategic.

The solution proposed in this research is to design a feedback (closed-loop) control system for cable networks using dynamic pricing. The goal of the control system is to make the regulation of cable networks automatic. The solution consists of three phases: modeling and analysis of cable networks, observer design of cable networks, and controller synthesis of cable networks. Each phase by itself tackles several challenges in the fields of computer networks, control theory, and mathematics. This research is intended to enhance bandwidth utilization efficiency, fairness, and security strength. The creation of an autonomic cable network involves two objectives:

1. Devise a mechanism (protocol) to ensure that individual benefit is maximized, as well as social efficiency and fairness using a dynamic pricing scheme.
2. Design a feedback (closed-loop) control system for cable networks.

The proposed feedback control system for cable networks has some desired properties, such as controllability, stability, sub-efficiency, voluntary participation, incentive compatibility, simplicity, and practicability. The solution is validated by not only theoretic studies but simulations.

PREFACE

In the classes I took years ago, I found that control theory, especially feedback control theory, has achieved amazing success in creating man-made autonomic systems (also called smart machines), such as unmanned aircraft, robots, and automatic assembly lines, and it is one of the most elegant theories so far. In the computer network performance analysis class I took in the same semester, a lot of performance analysis was talked about, but I had a question: Can we make a computer network, like a missile, act what we desire, not just accept its performance? This question motivated me to investigate the application of control theory principles to computer network control.

Control of computer networks solves their management problems. The OSI network management functions can be classified into five categories: fault, configuration, accounting, performance, and security. Also, there are many types of computer networks. What management function should I focus on, and what type of networks should I work on? I realized that resource allocation and quality-of-service are two of the primary concerns in a computer network, which usually consists of many autonomous participants. If these participants act strategically, the quality of service of the network may be unachievable. Then, I chose to study resource allocation and quality of service in cable networks, in which each subscriber acts independently.

Packet-switched networks provide the means to promote resource utilization and economic efficiency by means of resource sharing, compared with the dedicated communication channels of the traditional telephone. However, in addition to efficiency, predictable network quality of services is still desirable for many mission-critical applications, such as control over networks, online treading, and audio/video applications, which have strict temporal requirements. Nevertheless, it is unrealistic for a current open-loop computer network, without consideration of its participant autonomy, to provide predictable quality of service. By treating deviations from desired network quality of service as disturbances, feedback control which takes user autonomy into account offers a promising solution.

Considering its participants as autonomous agents, a computer network is a community. Because of limited network resources (scarcity in economic terms), contentions and externalities can exist. Thus, some economic principles can be applied. Game theory, which originated in economics, can be applied to predict the outcomes of these contentions, social welfare metrics can be used to measure the performance of this community, and a pricing mechanism can be utilized to regulate it.

This dissertation research provides only a starting point for predictable services and automation design/implementation with control theory principles, since it addresses a local network only. Typically, before flowing into the destination, a packet needs to pass several autonomous ISPs. So, predictability of network services is still questionable. For example, an ISP may favor the traffic flows of its own subscribers. There is a lot of work yet to be done.

This dissertation includes three parts, the modeling and analysis of cable networks, traffic classification and observer design, and controller synthesis. The contents of the modeling and analysis of cable networks and the traffic classification and observer design were summarized into two research papers, which were published in 2012 for an Annual IEEE Communications Quality and Reliability (CQR) International Workshop. Hopefully, this dissertation is able to stimulate fruitful discussion.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

Chapter Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF ABBREVIATIONS/NOMENCLATURE

AP	Access Point
ARMA	Auto-regressive Moving Average
BE	Best Effort
BWA	Broadband Wireless Access
CBR	Constant bit rate
CM	Cable Modem
CMTS	Cable Modem Termination System
DMS	Data mini-slots
DOCSIS	Data over Cable Service Interface Specification
DoS	Denial of Service
ESW	Egalitarian Social Welfare
EWMA	Exponential Weighted Moving Average
FTP	File Transfer Protocol
HFC	hybrid fiber coaxial
HMM	Hidden Markov Model
HTTP	Hypertext Transfer Protocol
IPS	Inter-packet time
ISP	Internet Service Provider
Kbps	Kilobytes per second
MAC	Media Access Control
MAP	Bandwidth Allocation Map
Mbps	Megabytes per second

LIST OF ABBREVIATIONS/NOMENCLATURE (continued)

MDP	Markov Decision Process
MPC	Model Predictive Control
ms	Millisecond
NE	Nash Equilibrium
NP	Non Polynomial
PID	Proportional Integral Derivative
Pkt	Packet or Frame depending on the context
PS	Packet Size
QoS	Quality of Service
RED	Random Early Detection
Req	Request
RMS	Request mini-slots
RTT	Round trip time
rtPS	Real-Time Polling Service
S-CDMA	Synchronous Code Division Multiple Access
SVM	Support Vector Machine
SW	Social Welfare
TAP	Transit Access Point
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
Tx	Transmit
UGS	Unsolicited Grand Service

LIST OF ABBREVIATIONS/NOMENCLATURE (continued)

USW	Utilitarian Social Welfare
VCG	Vickrey-Clark-Groves

LIST OF SYMBOLS

$\Re$	Set of real numbers
α	Probability that heavy CMs are idle
β	Probability that light CMs are idle
Δ	Difference between the back-off probabilities of heavy users and light users
Δ_j	j^{th} time interval of T_{delta}
$\Delta b_i^{(j)}$	Bit count of the fl_i in Δ_j
ε	Heavy-side function
θ	EWMA parameter
$\kappa_{i,k}$	Sensitivity of the transmission rate change with time to the utility change with the allocation for CM_i at time k
λ	Cable network HMM model
λ_1	Utilization HMM model
λ_2	Fairness HMM model
μ_i	Poisson process intensity of CM_i
$\hat{\mu}_i^{(j)}$	Estimated Poisson process intensity of CM_i in Δ_j
π	Initial state probability distribution of an HMM
π_i	i^{th} initial state probability of an HMM
σ	Sub-system index variable (σ = 1, 2)
A	$A = A_1 \times A_2 \times \ldots A_n$. The action space of a cable network

LIST OF SYMBOLS (continued)

AP1	Action profile: all CMs obey
AP2	Action profile: only one CM, CM_i, disobeys
AP3	Action profile: two or more CMs disobey
$A_{hmm} = \{a_{ij}\}$	State transition probability matrix of an HMM
A_i	A set of actions (strategies) of CM_i
Alloc	A vector of normalized bandwidth allocations, i.e. the state variables of all CMs
a_i	Action variable of CM_i
a_{-i}	Vector of the action variables of CMs excluding CM_i
$\hat{a}_i^{(j)}$	Estimated allocation of a cable network in Δ_j
a_{imax}	CM_i's action, which maximizes the utility of CM_i
$allocsum_{eq}$	Network utilization under the Nash equilibrium
$B_{hmm} = \{b_j(k)\}$	Observation symbol probability distribution conditional on the current state j of an HMM
CM_{cn}	A set of CMs
CM_i	i^{th} CM in CM_{cn}
CN	Multi-stage cable network model
C_{obs}	Cable network observer model
cap	Capacity of a cable network
c_i	Weight of i^{th} social welfare function for the linear aggregate social welfare function
D	Demand function, D: $p \rightarrow s$

LIST OF SYMBOLS (continued)

D^{-1}	Inverse demand function
E	System switching event set
E_c	A set of controllable cable network events
E_{uc}	A set of uncontrollable cable network events
F	$F = F_1 \times F_2 \times \ldots F_i \ldots \times F_n$, the flow feature space for a cable network
FAIRNESS	Network fairness space
FC	A set of flow classes
FL	A set of traffic flows of a cable network
F_i	Feature space for flow i corresponding to CM_i.
f	A flow classification assignment, i.e. a map f: FL $\rightarrow$ FC
$f_\sigma(y, p)$	Dynamic function of cable networks
fl_i	Traffic flow of CM_i
g	A network state assignment, i.e. a map g: NF$\rightarrow$NS
I	A set of interaction rules
IBoW	Initial back-off window
K_1	Sub-system 1 (σ = 1) controller gain
K_2	Sub-system 2 (σ= 2) controller gain
L	Number of social welfare functions
M	Cardinality of observation space of an HMM
MBoW	Maximum back-off window
max_retry	Maximum retry

LIST OF SYMBOLS (continued)

N	Cardinality of network hidden space
NF	A set of network features
NS	NS = (PHASE x FAIRNESS), the abstract state space of the cable network
n	Number of CMs in the cable network
n_1	Number of light CMs
n_2	Number of heavy CMs
o_t	HMM observation at time t
P	A set of cable network policies
PHASE	Network phase space
p	Policy variable in a cable network, or price depending on context
$p_{bj}(w_k)$	Back-off probability of CM_j with the back-off window, w_k
$p_{c\alpha}$	Collision probability of a heavy user
$p_{c\beta}$	Collision probability of a light user
p^d_{bj}	Back-off probability of CM_j when CM_i disobeys and the rest of CMs obey
p_h	Price users are willing to pay corresponding s_h
p_k or $p(t_k)$	Dynamic price at time k
p_l	Price users are willing to pay corresponding s_l
p_{max}	Policy that maximize network social welfare objective, or the maximum price users are willing to pay depending on context
p_n	Price users are willing to pay corresponding s_n

LIST OF SYMBOLS (continued)

p^{o}_{bj}	Back-off probability of CM_j when all CMs are obedient
p^{o}_{ti}	Request transmission probability of CM_i when all CMs are obedient
q_t	HMM hidden state at time t
r,	Interaction rule variable of a cable network, or the back-off random # drawn from a uniform distribution depending on context
S	$S = S_1 \times S_2 \times \ldots S_n$. The state space of a cable network, i.e. an allocation space
SW_{cn}	A set of nonnegative SW functions of a cable network
S_i	A state space of CM_i
s	Normalized utilization of a cable network
$\hat{s}$	Estimated utilization social welfare function
s_d	Difference between the normalized utilization and the provisioned normalized utilization
s_h	Normalized maximal bandwidth demand of the region corresponding to light use
s_{heq}	State of any heavy CM under the Nash equilibrium
s_i	Normalized bandwidth allocated to CM_i, i.e. the state variable of CM_i, or the hidden state of an HMM depending on the context
$s_{i[allobey]}$	State of CM_i when all CMs obey
$s_{i[disobey]}$	State of CM_i when CM_i disobeys and the remaining CMs obey
$\hat{s}_i^{(j)}$	Estimated state of CM_i in Δ_j
s_l	Normalized maximal bandwidth demand of the region corresponding to heavy use

LIST OF SYMBOLS (continued)

s_{leq}	State of any light CM under the Nash equilibrium
s_{max}	Maximum network utilization
s_n	Normalized maximal bandwidth demand of the region corresponding to normal use
s_{pre}	Prescribed state of a CM
s_t	Normalized utilization at time t
$\tilde{s}_t$	Smoothed utilization estimate at time t
s_{th}	Prescribed/provisioned network utilization
sw_{cn}	A real-valued aggregated social welfare
sw_{egal}	Egalitarian social welfare function
$\widehat{sw}_{egal}$	Estimated egalitarian social welfare function
sw_j	j^{th} social welfare function of a cable network
sw_{util}	Utilitarian social welfare function
$\widehat{sw}_{util}$	Estimated utilitarian social welfare function
T	A set of stages
T_{delta}	A set of small identical time intervals
th_l	Lower allocation threshold factor
th_h	Upper allocation threshold factor
U	A set of utility functions of CMs
u_i	Utility function of CM_i
$u_{i[allobey]}$	Utility of CM_i when all CMs obey

$u_{i[disobey]}$	Utility of CM_i when CM_i disobeys and the remaining CMs obey
V	A Lyapunov function
v_i	i^{th} 1-dimension HMM observation symbol or the valuation function of CM_i depending on context
w_k	Back-off window at back-off stage k
$x_{i,k}$	Transmission rate (bits/s) determined by CM_i at time k
y	Bandwidth allocation vector of a cable network
y_i	Bandwidth allocation of CM_i
y_{it}	Bandwidth allocation of CM_i at time t

CHAPTER 1

INTRODUCTION

1.1 Overview of Network Resource Allocation and Control

Resource allocation of computer networks is a universal problem. The goal is to allocate the resources of computer networks efficiently and fairly. Efficiency of resource allocation means two things: (1) high resource utilization; (2) resources are directed to where they are most needed, also referred to as economic efficiency. Media sharing is supposed to promote resource allocation efficiency but may impair it as well as fairness when the control scheme is inappropriate.

Control schemes in protocols such as media access control (MAC) and transmission control protocol (TCP) congestion control have been used to resolve conflicts among agents sharing network bandwidth. Data over Cable Service Interface Specification (DOCSIS) for cable networks is an example of MAC. However, these control schemes assume that network participants adhere to the specifications and perform the prescribed actions. These protocols do not account for the possibility of autonomous network participants deviating from the prescribed actions in an attempt to maximize their benefit. A common deviation is heavy use of bandwidth which may result in network congestion and unfairness. From the view of control theory, current cable networks are primarily open-loop systems. Treating departures from prescribed actions as disturbances could make feedback control a feasible solution. Other control schemes to avoid congestion include pre-allocating a fixed amount of bandwidth to a specific user statically. However, these policies lead to under-utilization of bandwidth and low social welfare.

A computer network where resources are shared by multiple participants is a multi-agent system. In such a system, participants compete for a fixed amount of resources which usually leads to externality (the actions by one group of agents affecting the welfare of others). Pricing is a common mechanism to regulate competition for the limited resources. Flat pricing is the simplest pricing scheme. This scheme is often successful in companies which reward the efficiency of workers by paying per task instead of by the hour. However, flat pricing creates resource allocation problems in cable networks. There was the similar problem with flat pricing in the early-year distribution of electricity.

In addition to the above discovered problem in cable networks, there is another potential problem that, at least in theory, network participants are able to implement protocols which deviate from standard protocols in an attempt to increase personal benefit. Since the deviations are related to denial-of-service (DoS) attacks, they are a network security issue. Thus, resolving these allocation problems not only enhances the performance of a computer network but promotes its security.

1.2 Dissertation Overview

This research proposes a feedback control system to solve the allocation issues for cable networks. A feedback control system signals the plant to correct deviations when a difference is recognized between the actual system performance and the predefined one. An example of feedback control systems is the missile control system, which ensures that the missile hits to the target within some acceptable accuracy even though it encounters unpredictable strong wind along its way. Control theory, especially feedback control theory, has achieved great success and revolutionized the world in the past forty years in automated systems such as missiles, unmanned aircrafts, robots, automatic assembly lines, and others. It is also widely used to study the self-regulation of biological systems, such as the respiratory system.

In recent years, control theory has been recognized as a viable technology for enabling the construction of autonomic computer networks as well. However, control of networks is still a young field, and significant problems with applying control theory to real computer networks still exist. The control of computer networks is difficult because of their great complexity and large scale. The challenges in both theory and application are substantial, especially in the areas of network modeling, observer design, and controller synthesis.

An effective control system design requires a model of the system behavior. This research relies on a model for both the controller design as well as the evaluation of cable networks. Game theory is often used to model multi-agent systems where agents are strategic and externalities exist. The most popular solution concepts of games are Nash equilibriums (NEs). The NE concept is widely used to predict the outcome of the strategic interactions among decision makers in economics, political science, and other sciences. When the assumption of non-strategic participation is relaxed, a cable network is a multi-agent system where the actions of one participant affect the welfare of others. This research regards a cable network to be a community of agents and applies game theory, combined with system theory, to model the cable networks.

Social welfare functions can be used to quantify the overall welfare of society. When applied to income distribution, social welfare functions evaluate the equality of wages and the total wages of the community. The most common social welfare measurements are utilitarian social welfare (the sum of individual utilities) and egalitarian social welfare (an agent's utility which is the worst off). Social welfare metrics can measure both the efficiency and fairness of a society, allowing the tradeoffs of these factors to be measured for various situations. In this research, the cable network is considered the community, and the cable modems are the agents.

An agent's utility is defined as a convex function of the bandwidth allocated to it. Using social welfare metrics to evaluate cable networks allows them to be evaluated not only in terms of resource utilization but also in terms of fairness. Network utilization, which plays an important role in this research, is a special case of the utilitarian social welfare.

An observer is needed for a control system if a system state required by a controller is not directly observable. For example, assume it is needed to design a control system based on the velocity of a car inside a tunnel. The entering and exit velocities are directly observable by a satellite, but the velocities inside the tunnel can only be estimated. A more complex example is the underlying drives of stock indices, such as the psychological influences on investors. The stock prices are directly observable, but the investors' psychological factors can only be estimated. The metrics of these underlying factors can be thought of as abstract states. The observer proposed in this research classifies the network flows and the network states, which are not directly observable, and provides these as inputs to the controller. The controller can then use the current information about the flows and state of the network to provide control signals to the system.

Economic theory teaches that the prices of resources can successfully control their level of use. For example, charging a higher fee for use of a public park reduces the competition for its space and free access promotes its usage. In the same way, pricing can control the utilization and fairness of cable networks. In this research, a new dynamic rather than static pricing scheme will be used to control the frame sending rates of the subscribers of cable networks in order to promote network utilization and fairness. Therefore, the cable network being controlled will be modeled as a dynamic game, in which the current strategy of an agent depends on the previous actions selected. A controller will fulfill the objective of this resource allocation problem. The controller obtains the state information and classification results, and then generates control signal to the cable networks. This multi-agent control system will ensure that every participant gains and social efficiency is maximized under fairness constraints.

1.3 Three Phases of Dissertation Research

The solution is simplified to three phases to tackle the complexity of this research: modeling and analysis of cable networks, observer design of cable networks, and controller synthesis of cable networks. First, feedback control of cable networks requires a system model. In this research, a system model was developed to study cable network performance when its participants function strategically. The model is based on game theory and system theory. The cable network studied is considered a multi-agent system, and the network quality of service is evaluated with several social welfare metrics. Theoretical studies concluded that the prescribed quality of service may be unattainable when the network subscribers act strategically, and that a better control mechanism is needed to encourage the network participants to realize the prescribed quality of service. In addition to the performance analysis, the model will be used for controller synthesis.

Second, in order to design a controller for a cable network, an observer is needed to evaluate the current network state and traffic flows. In this research, an observer model and a detailed observer design of cable networks was proposed. Separation of the observer model and its design is for the purpose of flexibility of implementation. The observer design is recast into an online traffic classification problem, which requires the identification of hidden states. This research constructs a hidden state space using social welfare metrics and the four-phases of network utilization dynamics for cable networks. An observer is then designed that classifies the network flows as well as the network state. The network state is classified using a hidden Markov model.

Third, the last phase of this research involves controller synthesis. The basic idea of controller synthesis is to design a dynamic game with pricing means so that maximization of the utility of each individual participant leads to maximization of social efficiency under a fairness constraint. Mechanism design was used to design the controllers. To demonstrate the effectiveness of the control system for cable networks, the properties, such as controllability, stability, incentive compatibility, voluntary participation, and sub-efficiency of the cable network with the proposed control system were studied.

1.4 Dissertation Outline

The remaining of this dissertation is organized as follows. Chapter 2 gives an overview of cable network and DOCSIS. Chapter 3 provides the literature survey on cable networks, multi-agent systems, game theory, network dynamic patterns, traffic classification, autonomic networking with control theory, system modeling techniques, observer construction methodologies, controller synthesis approaches, mechanism design, and dynamic pricing. Modeling of cable networks and construction of a cable network control system, including observer and controller, are presented in Chapter 4. In Chapter 5, theoretic analysis, simulation study of cable networks, and evaluation of the control system are provided. Chapter 6 concludes the dissertation and lays out some future work.

CHAPTER 2

OVERVIEW OF CABLE NETWORKS AND DOCSIS

A cable network is a broadband access network, and Data over Cable Service Interface Specification (DOCSIS) is a level 1-2 protocol for it. The DOCSIS specification is defined in [1]. A conceptual cable network is shown in Figure 1. A cable network consists of several Cable Modems (CMs) and a Cable Modem Termination System (CMTS). The CMs connect to the CMTS through either all-coaxial or hybrid fiber coaxial (HFC) cables using a treelike architecture. The CMTS connects to the Internet through an operator's back office network, and each CM attaches to a subscriber's network. Cable communication is divided into several channels, which have different frequency ranges. Data channels are categorized into two groups: upstream transmission data channels (from CMs to CMTS) and downstream transmission data channels (from CMTS to CMs).

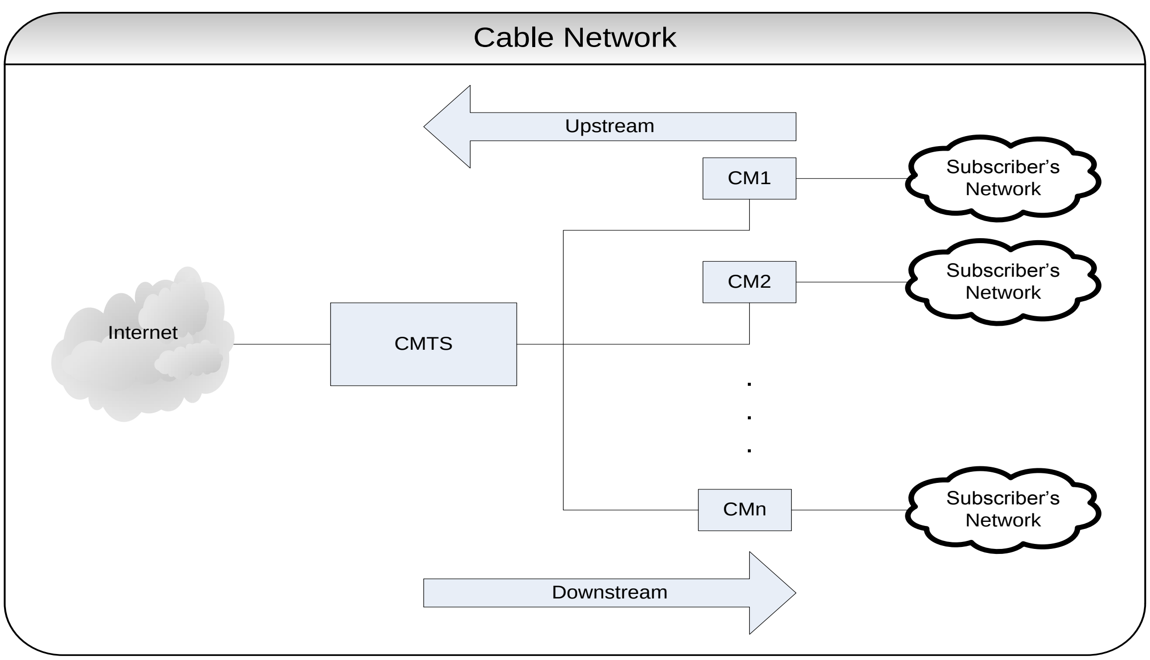

Figure 1. A cable network

DOCSIS has evolved from DOCSIS 1.0, DOCSIS 1.1, and DOCSIS 2.0 to DOCSIS 3.0 since its initiation. A data channel supports either Time Division Multiple Access (TDMA) or Synchronous Code Division Multiple Access (S-CDMA). DOCSIS uses tiny time intervals called mini-slots for upstream data transmission and utilizes Reservation Aloha protocol to schedule bandwidth allocation. To support different service flows with different Quality of Service (QoS) requirements for upstream data transmission, DOCSIS defines a variety of service types, including Unsolicited Grand Service (UGS), Real-Time Polling Service (rtPS), and Best Effort (BE) Service, with diverse QoS parameters. Each cable network upstream traffic flow is mapped to one of them.

BE service uses a request policy which permits CMs to make use of contention request opportunities for upstream transmission. Thus, collisions may occur during request transmission, and CMTS decides which mini-slots could be subject to collisions by sending Bandwidth Allocation Maps (or just called MAPs) to CMs. The format of MAP is shown in Figure 2. The collision resolution scheme, which is similar to that in Ethernet networks, is based on a truncated binary exponential back-off with two parameters (the initial back-off window and maximum back-off window) determined by CMTS. The back-off process uses uniform distribution. Figure 3 shows a flow chart for the collision resolution process with the multiple transmit channel mode disabled.

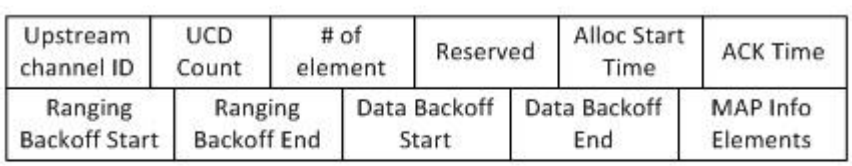

Figure 2. MAP format

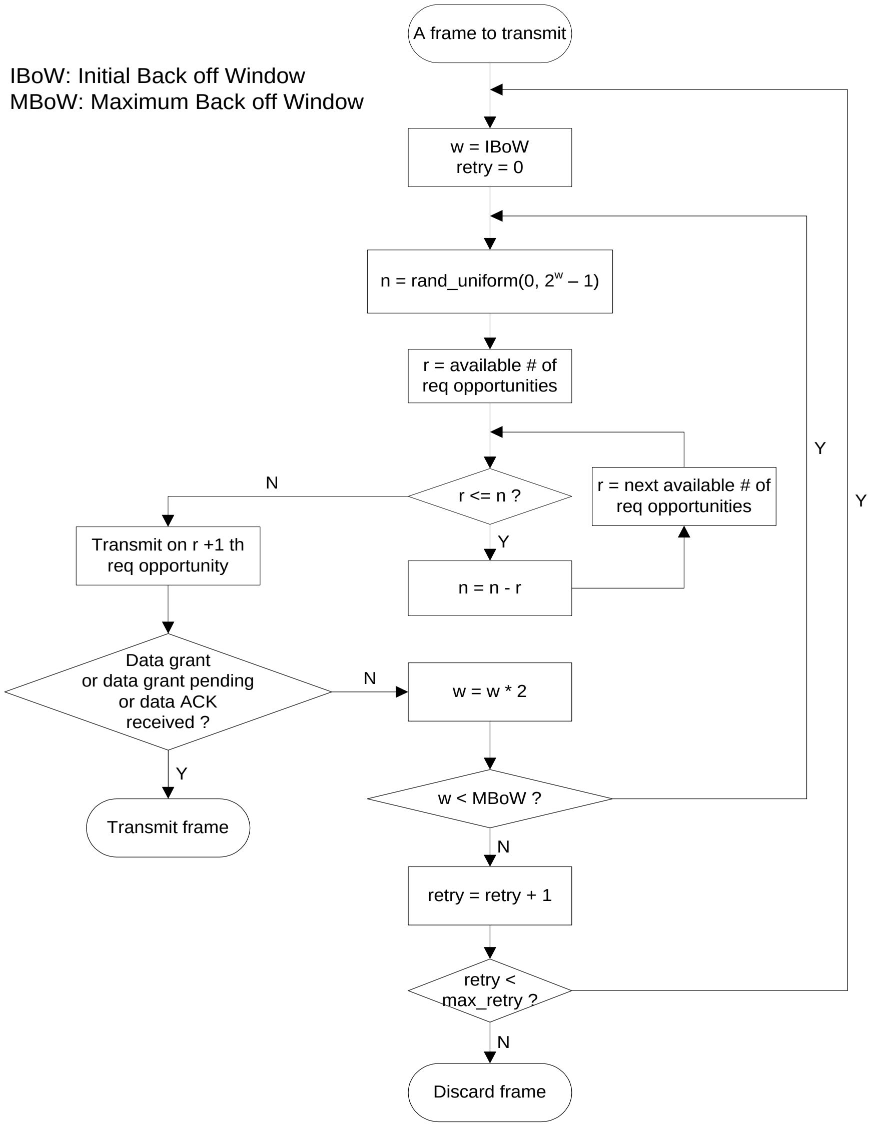

Figure 3. DOCSIS collision resolution process

CHAPTER 3

LITERATURE SURVEY

3.1 Introduction

This dissertation research is multi-disciplinary in nature. Literature survey on several fields, such as multi-agent systems and game theory, cable networks, network dynamic pattern recognition, traffic classification, autonomic networking with control theory, and mechanism design and dynamic pricing were conducted. A summary is given with regard to the properties of computer networks, the problems with cable networks, and the problems with current solutions to control of networks.

3.2 Multi-agent Systems & Game Theory

Multi-agent systems have been studied for decades, and multi-agent system modeling and control have become a growing area of research. A multi-agent system, such as a computer network, usually exhibits the property of emergency, which cannot be seen at the agent level. Due to lack of invariance principles in multi-agent systems, modeling multi-agent systems presents a significant challenge.

Game theoretic models are used to model the multi-agent systems where the agents are strategic and externalities exist [2] [3] [4]. The most popular solution concepts of games are Nash equilibriums (NEs). The NE concept is widely used to predict the outcome of the strategic interactions among decision makers in economics, political science, and other sciences. Recently, it also is used in some sub-fields of computers networks, such as wireless networks [5] [6]. The model proposed in this research was built on game theory.

Metrics to measure the performance of multi-agent systems is a critical element of multi-agent systems. One type of metrics used is social welfare metrics [7]. The social welfare metrics include utilitarian social welfare (the sum of individual utilities), egalitarian social welfare (an agent's utility which is the worst off), and Nash product (the product of individual utilities). In this research, a cable network is viewed as a community, the well-being of which is evaluated using social welfare metrics.

3.3 Cable Networks

The performance issues, such as link throughput and frame delay, of cable networks are investigated in [8] [9] [10]. The analytical model in [9] assumed that the cable network was under heavy traffic and that CMs were non-strategic. This research continues the assumption that the cable network is under heavy traffic in certain scenarios to simplify the analysis, although other assumptions such as a non-strategic CM are lifted in this work.

The problems with subsidization and low bandwidth utilization have been analyzed using datasets of commercial and experimental broadband networks in [11] [12]. A statistical analysis of a commercial broadband usage dataset for a Lite service in May 2002 reveals that the top 10% of subscribers produced 54.6% of the entire traffic whereas the bottom 10% generated only 0.3% [11]. That is, the top 10% consumes 182 times as many resources as the bottom 10%. It also shows that, if the monthly per-user flat usage fee the Internet Service Provider (ISP) had charged is assumed to be $10 without profit, a subscriber in the top 10% gains $44.6 per month while one in the bottom 10% pays $9.7 in excess per month [11]. An analysis of an experimental broadband dataset also gives evidence that the flat-rate scheme requires light subscribers to subsidize heavy ones and wastes network resources by causing low bandwidth utilization [12].

3.4 Network Dynamic Patterns & Traffic Classification

It is reported that some network traffic exhibits four phases: free flow, saturated, hysteretic (bi-stable), and jammed (congested) [13]. Phase transitions between free flow and congestion are also observed in the empirical study of computer networks [14] [15] [16] [17] [18]. This research builds on this observation and incorporates the traffic phase into the network state classification.

A traffic classification problem contains three parts: flow sampling, feature extraction, and traffic classifications. Each of these components has been studied extensively. Most sampling techniques target the problems of limited memory size, processing speed, and reporting bandwidth because of such huge volumes of network traffic [19] [20]. Some examples of flow features extracted include packet count, byte count, and bit rate [19] [20]. In this research, the bit count of each traffic flow of a cable network is estimated depending on the method in [19]. This method uses sampling (collecting a fraction of packets from a flow with some probability to reduce processing overheads) and slicing (tracing a slice of a flow with some probability to decrease memory usage) so as to be scalable to high speed networks. It also uses a reporting threshold to control the data volume to be reported. The rate of each traffic flow of cable networks is determined by one of the flow estimations proposed in [20]. The estimation makes use of individual flow bit measurements at two successive time points to compute the average bit rate of the flow in this time interval.

Machine-learning-based network traffic flow classification, which uses the statistical characteristics of packets, such as inter-arrival time, idle time, and packet length as the feature space, has been investigated intensively recently to overcome the limitations of port-based and payload-based network traffic flow classification techniques [21] [22] [23] [24] [25] [26] [27]. Flow classes can be broad classes, such as interactive, bulk, and transactional, or protocol families, such as HTTP, Telnet, and FTP. One of the machine-learning-based techniques used for the traffic flow classification is hidden Markov model (HMM) [22] [26] [27]. The key features used in these HMMs are packet size (PS) and inter-packet time (IPS). In [22], the feature vectors are quantized with k-means clustering to map a multi-dimensional space into a 1-dimensional discrete space. In addition to these supervised flow classifications, clustering and entropy-based techniques are used to identify flows “standing out” from the rest [23]. Basic concepts of HMM and other machine-learning-based techniques, such as support vector machine (SVM), can be found in [28] [29].

To satisfy the real-time requirement, online network traffic classification is also researched [24] [25]. In [24], an online traffic classification is presented to capture concept-drift dynamics in data stream. Concept drift means that the statistical properties of the predictable variable unforeseeably change over time [24]. A naïve Bayes classifier based on statistical features of sub-flows of 5–8 packets is used for an online traffic classification in [25]. Petri nets are also used for online classification problems such as event recognition in surveillance video [30]. An online traffic classifier with HMM is built for observer construction in this dissertation.

3.5 Autonomic Networking with Control Theory

Control theory directly or in principle has been applied to autonomic computing applications, and significant advantages of this approach have been demonstrated. In the literature, three terminologies, control of networks, intelligent networking, and autonomic networking, are used interchangeably since control theory is a major approach for autonomic networking. Three major aspects of control theory, system modeling techniques, observer construction methodologies, and controller synthesis approaches, are surveyed one by one.

3.5.1 System Modeling Techniques

Like traditional control systems, control of computer networks requires models (including process models) of computer networks. System modeling is one of the key ingredients to construct a control system. Selecting an appropriate model can simplify the design and implementation of control systems, and, at the same time, can improve the effectiveness of control.

The major models of computer network systems are classified in Figure 4 [31]. Based on the assumptions made about a computer network, the models range from flow-based models, discrete event models (including automata-based models and Petri net based models), and hybrid models (combination of flow-based models and discrete event models) to statistical models and heuristic models. Each of them has advantages and drawbacks. A flow-based system model is used for congestion control over the Internet [32]. Markov models are applied to a host-based autonomic defense system in [33]. Pattern based models (examples of statistical models) are used in [34] [35]. Petri net models together with fuzzy models are used for a network flow control [36]. TCP flow control with feedback mechanisms with a hybrid model is presented in [37].

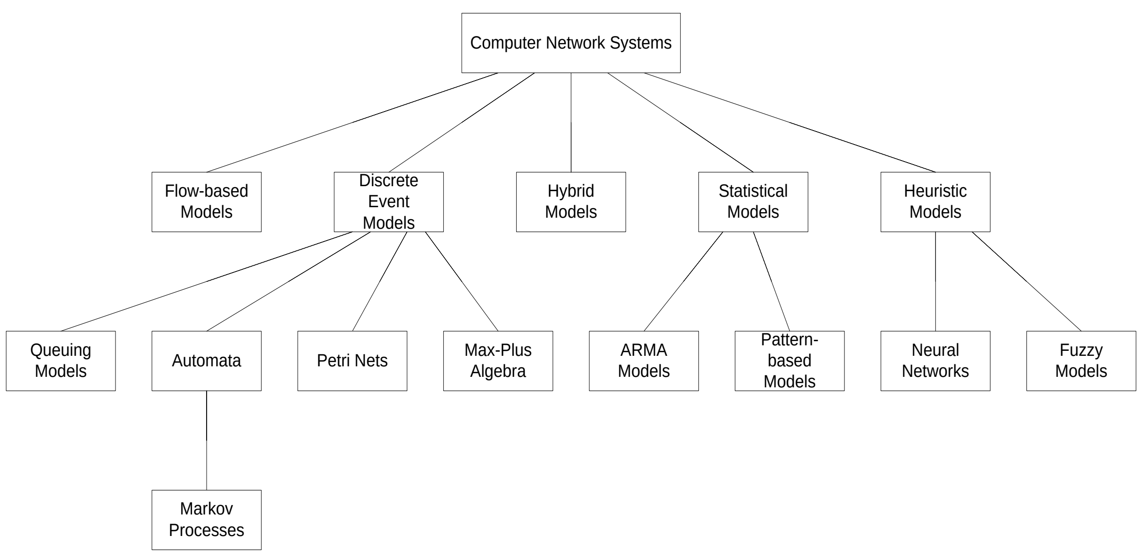

Figure 4. Model classes for computer network systems

A system modeled with a hybrid model, which is of great interest, is also called a switched system in control-theory literature. One advantage of switched system control is of dealing with model uncertainties, including un-modeled dynamics and model parameter variations. [38] Switched system control can be thought of a variation of adaptive control. [38] In this dissertation, a cable network is treated as a switched system by separating state space into two regions. Control signals are different in these two regions.

3.5.2 Observer Construction Methodologies

Observer construction has not been a great focus of most research on control of computer networks. In [32], the control system does not use an observer since the controller directly makes use of the output from the plant. In [33] [34] [39], each reported control system has a component similar to an observer, but no design information is described except that it is reported in [33] that the state estimator is recursive, which is a key characteristic of an observer. In [40], QoS provisioning for a Web server is proposed and the control system uses a simple linear filter as an observer to estimate server utilization based on two observable variables: server request rate and the delivered byte bandwidth. The coefficients of this filter are determined by a linear regression.

A family of low-pass filters, which do not require system dynamics to be known, is the moving-average family [41] [42]. The most popular member of this family is exponential weighted moving average (EWMA). EWMA was used for round trip time (RTT) estimate [41]. The EWMA parameter is suggested to be 7/8 in [41] and 3/5 in [42]. A formal modular model, HMM, and EWMA are utilized to construct an observer in this dissertation.

3.5.3 Controller Synthesis Approaches

It appears that most traditional control approaches have been investigated to address autonomic networking problems. Control approaches can be classified by control strategies, controller intelligence levels, controller organizations, and control policies, as shown in Figure 5 [31]. A fuzzy logic controller (belong to the category of intelligent control) is used for congestion control over the Internet [32]. Markov Decision Process (MDP), which uses the principle of model predictive control (MPC), is applied to a host-based autonomic defense system in [33]. An adaptive proportional-integral-derivative (PID) controller based on pattern recognition to achieve adaptive control for guaranteeing the desired QoS of computer networks is reported [35]. Feedback control mechanisms are used for TCP flow control [37] and QoS provisioning for a Web server [40].

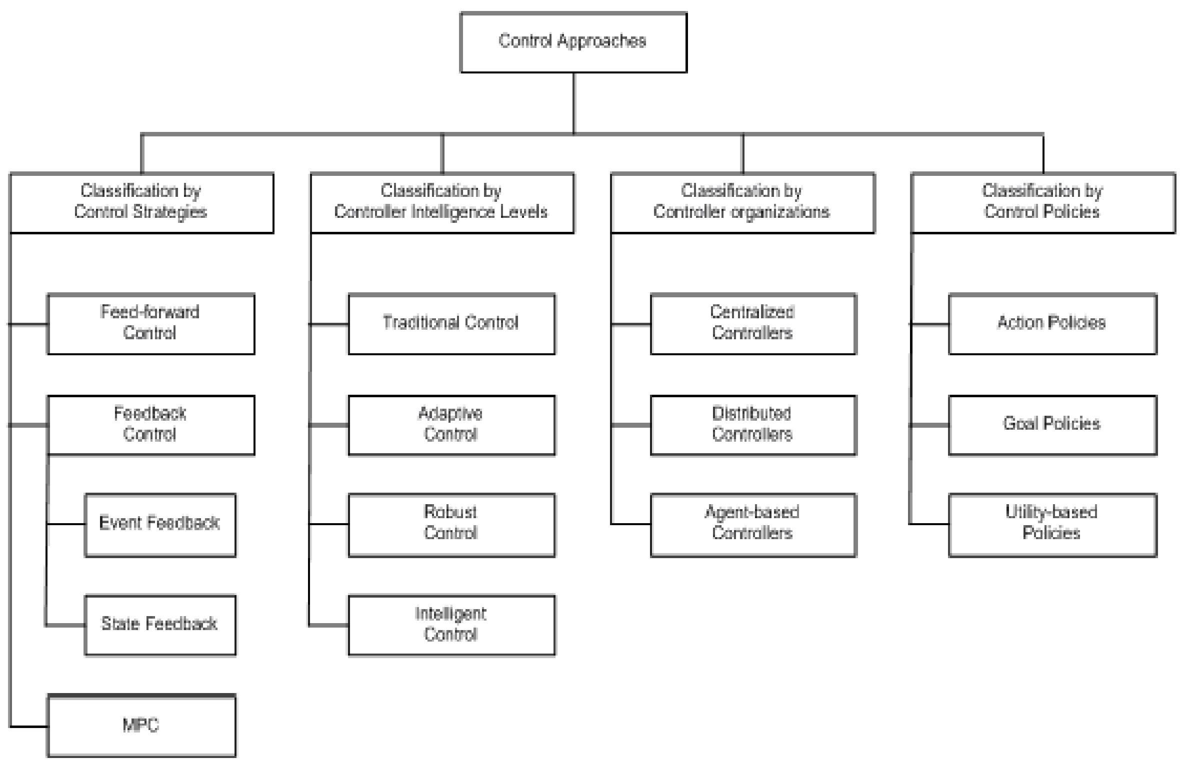

Figure 5. Classification of control approaches

Sliding mode control is a non-linear control strategy, in which the feedback control law is not continuous in time [43]. Sliding mode control can handle model uncertainties effectively [43]. In this dissertation, sliding mode control together with switched system control is used to cope with non-linearity and model uncertainties of cable network systems.

3.6 Mechanism Design & Dynamic Pricing

Mechanism design is a methodology to design a new game with desired properties such as incentive compatibility and efficiency [44] [45] [46]. Mechanism design originates from economics. Mechanism design determines two rules: allocation rule and payment rule. Allocation rule determines how to allocate resources to agents and payment rule set the prices for the resources allocated. Since payment rule is usually dynamic, payment rule often refers to dynamic pricing.

Mechanism design has been applied to design protocols of computer networks, typically wireless networks. The design is either through an existent mechanism (such as Vickrey-Clark-Groves (VCG) mechanism) [47] [48] [49] [50] [51], which has been proved to have some desired properties, an optimization [52] [53] [54], which requires accuracy of system models and usually has non polynomial (NP) hard computation complexity, or a heuristic [55] [56], which trades off efficiency for simplicity. However, these design methodologies either require that the agents' valuation functions are known to the designer or that agents reveal their valuation functions to the designer. In reality, these individual valuation functions are hardly known and revelation incurs large communication overheads. This dissertation relaxes this assumption and follows a heuristic paradigm.

Control, which does not sacrifice the best-effort nature of packet-switched networks, for protecting best effort service is promoted [54] [57]. "Indeed, extending the traditional best effort service model has proven to be not only an extremely complex task but quite controversial too." [57] Current congestion control approaches can be categorized into three classes: per-flow scheduling (such as admission control and resource reservation), penalty box (e.g. dropping elephant flows), and pricing [57] [58]. Incentive compatibility is advocated rather than excluding non-responsible peers [51]. Only the pricing method, which can make a control incentive compatible, preserves the nature of the best effort. Thus, this dissertation follows the pricing approach.

The idea of using dynamic pricing in computer networks is not new. The earliest paper found about dynamic pricing is published in 1993, which investigates user incentives in multi-service-class computer networks with best effort services [59]. It is argued that with service-class insensitive pricing, i.e. the flat pricing, network efficiency can be very low. Therefore, a service-class sensitive pricing scheme is needed to achieve the desired level of network quality of service. A simulation shows that, under service-class sensitive pricing, the users are satisfied more and optimize network system efficiency on the whole while maximizing their own utilities. Another interesting finding is that a pricing scheme is less influential while the network has light load. However, flows are categorized by application types and there is no dynamical treatment of these flows. In addition, technical details and formal proofs are not presented.

A price-based congestion control scheme for service differentiation in broadband wireless access (BWA) networks is presented [60]. "While dynamic pricing is a powerful tool for controlling congestion in BWA networks, it has received little attention and is still an open research issue." [60] Two levels of games, the bandwidth provisioning at the class level and the scheduling at the connection level are proposed. However, there is lack of the details regarding determination of the price factor at the class level and, also, there is no connection level pricing found.

This dissertation also uses the idea of dynamic pricing. However, in this dissertation, a formal detailed technical solution with control theory principles is proposed for both determination of pricing and automation of network devices at connection level. The dynamics of flows, which are abstracted with three broad classes, heavy, normal, and light, are handled in this dissertation.

A dynamic network bandwidth allocation approach based on microeconomic principles is discussed to show the effect of pricing in communication networks. In [61], QoS profiles are imbedded in users' utilities. In each time period, the prices of links are adjusted dynamically by the demands for bandwidth (supplies are fixed) and an amount of bandwidth purchased on a link is determined according to a price set for this link. A network broker polices the users to ensure that the sending rate is less than or equal to the bandwidth bought on a link. A simulation study is provided to show that this approach achieves high link utilization and high fairness. However, no formal proofs of these properties are provided. In addition, a disadvantage of this approach is that detection of bandwidth demand and determination of price in each time period can cause considerable user decision-making, communication, transaction, and computation overheads even though the network load is light. This approach is not appropriate for media sharing, in which users' sending rates are unknown for a social agency.

Dynamic pricing scheme for call admission control for wireless networks is proposed to alleviate congestion [62]. As in [61], the price is adapted to the network load condition. A normal price is charged for each user when the arrival call rate is low and a peak rate is charged to the users who want to place calls at this network condition when the arrival call rate is high. Utility functions and demand functions for all users are assumed homogenous and known. The social welfare is measured with a linear combination of new-call blocking and handoff-call blocking probabilities. A pricing function block is applied to the system according to the current traffic condition to prevent it from congestion. The price is determined according to the optimal arrival rate, new arrival rate, and the delayed arrival rate. How to generate the delayed traffic stream from the original traffic stream (one of the key ideas) is not found. In addition, like the approach in [61], this mechanism is not suitable for media sharing since users' sending rates are unknown for a social agency. A linear regression estimation of demand curves during periods of a day for telecommunication traffic is studied. [63]

The work in this dissertation also uses microeconomics principle to reshape the traffic pattern to achieve the objectives of control of cable networks. The detailed engineering designs, such as state detection, price determination and device automation, are given. We relax the assumption of homogeneity of users and knowingness of user utility functions to reflect the reality. In this dissertation, formal treatments, such as modular design and rigorous proofs, are provided. Our control design aims for reducing the overheads of user decision-making, communication, transaction, and computation.

Money transfer could be unnatural, cumbersome or unrealistic because of transaction and communication overheads. "It is challenging to determine what such payments could be and how, in practice, they could be enforced." [64] It is found that, in some situations, network efficiency improvement can be achieved without an actual payment enforced while not as effective as a payment-base mechanism [64]. Some mechanisms set price but do not require a payment if agents act responsibly [64] [65] [66] [67]. A price actually acts as a warning. For example, a congestion fee is charged when the network under congestion and agents will not be charged if they comply with the social agency. This dissertation follows this concept in order to reduce the overheads of cable network operation modification and disruption.

Some mechanisms count on indirect monetary costs (such as collision rate and power consumption), resource rewarding, and virtual money. In [65] [67] [68], collision rate or power consumption experienced by agents themselves is treated as costs. An advantage of this treatment is that no central agency is required. Nevertheless, this kind of costs may be ignored by agents still resulting in mechanism failure. IEEE 802.11e networks studied are assumed to have two MAC-layer access phases, contention and pooling (contention free), and the polling phase is used to reward the behaviors of wireless nodes [69]. It is assumed that there are two types of traffic, low-priority and high-priority. Access point (AP) resorts polling to incentivize users to transmit their low-priority traffic with the right access probability. Since use of contention-free phase consumes system resource, this mechanism incurs efficiency loss in every bandwidth allocation configuration. Determination of pooling proportion is another issue too and a small portion of pooling may make the mechanism incentive incompatible. A mechanism for realizing fairness in multi-hop wireless backhaul networks is studied [70]. This mechanism uses two types of virtual money, credits and tokens. Each Transit Access Point (TAP) in the network earns credits by forwarding packets and consumes them by sending packets for its local users. A TAP needs to have enough accumulated credits to earn a token. How to determine the credits and the tokens is a complicated problem. A central bank is required for this mechanism to work, which results in additional overheads. This dissertation uses direct monetary charge as a threat.

To implement the optimum of a network rate control distributively, feedback control can be used by specifying a user's rate change per unit of time to be proportional to the difference between the cost a user is willing to pay and the cost charged [71] [72]. Gradient method is commonly used in traditional optimal design [73]. Gradient play is a mainstay of the update schemes of agents' sending rates [53] [65] [74]. Compared with best-response-based play or dominant-strategy-based play, gradient play does not concern about other agents' strategies and only the price set by the social agency matters. In this dissertation, based on what a user is willing to pay and the bandwidth cost charged, each CM uses gradient play to maximize its own utility and the network utilization at the same time.

Fairness is another important factor for resource allocation besides congestion control. In communication networks, fairness is closely related to congestion. “Fairness becomes an issue only when there are unsatisfied demands and users have to compete for their share.” [57] This is the sense of proportional fairness. There are other senses of fairness, such as max-min fairness [52] [71] [75]. This dissertation only considers proportional fairness at this time.

Pareto efficiency is one of properties mechanism designer wants to achieve. In a Pareto efficient allocation, nobody can become better off without harming another. The shape of Pareto efficiency frontier in terms of throughput and power consumption for wireless networks with two nodes is presented in [56] [76]. This dissertation makes use of a similar idea.

“Although game theory provides solutions for a situation of conflict during multiple access, it is still difficult to implement these solutions in a practical environment.” [77] One difficulty is from uncertainty in the beliefs of the agents [78]. This uncertainty results in system uncertainty. Combination of mechanism design and control theory provides a promising solution. Only a few literatures treat game design from the perspective of control theory [78] [79] [80]. Relevance of control theory to mechanism design is discussed for an auction design [78]. Deterministic multi-stage dynamic Stackelberg games are treated from a perspective of control theory [79]. Non-cooperative game design for an ad hoc wireless network is treated from the perspective of control theory [80]. It is assumed that the cost structure of agent i is of a quasi-linear function. Valuation functions of all agents are assumed known.

In this dissertation, game design is treated from the perspective of both mechanism design and control theory to provide a viable control solution for cable networks. The valuation functions of agents are assumed unknown and the social agency only has some idea of the market demand of a cable network. Switched control and sliding mode control are used to cope with the model uncertainties in cable networks.

3.7 Rate Control & Network Security

One of the ideas to prevent network attacks is to make attacking more difficult by limiting resource consumption. “A method for limiting resource consumption is the use of client puzzles.” [81] In the same vein, charging a fee for heavy use of cable networks can also prevent attacks by making attacks much more costly. Thus, this scheme improves not only network efficiency but also security.

3.8 Properties of Computer Networks

The following are the major properties of computer networks, including cable networks:

- Composition of autonomous systems, including independent users: A computer network composes of many autonomous systems/users, and the Internet or a cable network is such a network. These independent entities are called agents, which act independently with their own interests and policies.
- Best effort: Computer networks are packet-switched in nature and usually packets travel through several autonomous systems. Thus, unlike traditional circuit-switched networks, the delivery of packets can never be precisely guaranteed.
- Externality: A user’s behaviors may affect the QoS of others.
- Dynamic change: Network topology, network participants, and network conditions change over time.
- Large scale and distributiveness: Computer networks are distributed and of large scale by nature. Therefore, obtaining the global states of a computer network is virtually impossible.

- Discrete event systems: A computer network is a discrete event system, which is driven by discrete events instead of time.
- Nonlinear systems

3.9 Problems with Current Cable Networks

Computer networks have far reaching benefits. However, the following problems exist in current cable networks:

- Economic unfairness: A flat-rate pricing scheme is prevalent today. This pricing scheme causes light users to subsidize the heavy users and encourages users to overuse the network resources.
- Resources inefficiency (Low utilization of resources): Either resources are underused or performance is deteriorated by overusing the network resources (such as traffic congestion) due to self-interested cable network subscribers.
- Social/Economic inefficiency: Crucial traffic flows may not be handled in a timely manner. The packet delay or packet loss of these flows may lead to social or economic losses.
- Feed forward control mechanisms: Current control schemes behave like feed forward control mechanisms. Cable networks are basically open systems, and resources in cable networks are typically allocated by assuming every network participant acts as prescribed.

3.10 Problems with Current Solutions to Control of Networks

Significant progress in the control of computer networks has been made through substantial research efforts. However, the current solutions to control of computer networks still have the following issues:

- The flat-rate policy is dominating, probably because the previously proposed rate control solutions are difficult to implement. The implementation difficulty is due to inappropriate system models, measurement problems, and/or ineffective control approaches.
- It is assumed that the network participants are either obedient or adversary. An example is random early detection (RED), which assumes that each participant follows predefined rules. However, in reality, this behavior should not be expected since a user who does not follow the predefined rules may get a huge benefit.
- It is assumed the network global state is known by expecting all participants to reveal their true state. In practice, either most participants' states are hidden or an agent may benefit by not telling the truth even though this participant agrees to reveal his/her own states.
- Many solutions are point solutions which only address a single component of a network, such as a host, a router, or a server. Some solutions use the agent-based control approach, but they do not relax the above assumptions. In addition, scalability still remains a big issue.
- A large portion of research on the control of networks only discusses the framework of autonomic networking. The network modeling, detailed design and implementation are the most difficult issues which need investigated.

CHAPTER 4

FEEDBACK CONTROL SYSTEM FOR CABLE NETWORKS

4.1 Introduction

The solution proposed for this research is to design a feedback control system for cable networks, shown in Figure 6. The research was performed in three parts: modeling cable networks, observer design of cable networks, and controller synthesis of cable networks.

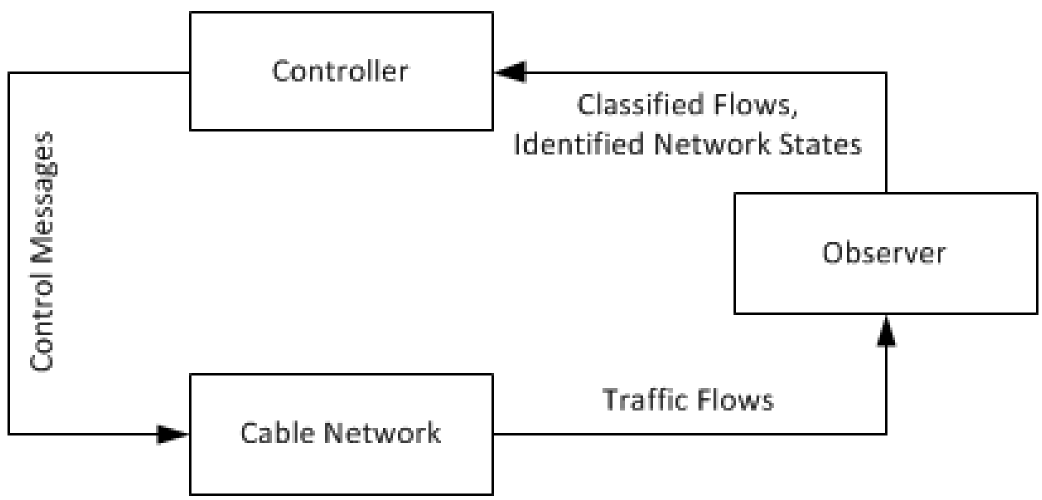

Figure 6. Feedback control of cable networks

The modeling of cable networks has three main purposes: determination of the model structure, definition of the state space, and description of the behaviors of the agents, the network, and the communication. The purpose of an observer is to classify network flows and to estimate network system states. The observer construction is recast into a classification problem. The vector of the micro states (also called bandwidth allocation) and the social welfare functions are the features of the classifiers. Controller synthesis is devoted to devising a mechanism (protocol) to ensure that the actions of maximization of individual benefit induce maximization of social efficiency under fairness constraint. The methodology of heuristic mechanism design is utilized to design a controller.

4.2 Assumptions, Conditions, and Limitations

The following are the assumptions, conditions, and limitations for this dissertation:

- Each agent, i.e. network participant, is rational. This means that each agent tries to maximize one's benefit.
- The same type of agents uses the same set of strategies.
- Each agent does not know the global states of the network in which it participates.
- The social agency does not know the individual bandwidth valuations of agents but the demand function of the cable network
- The scenarios analyzed are based on one ISP and several subscribers.
- The ISP acts as the social agency of the cable network.
- The number of subscriptions is fixed.
- There are at least a couple of heavy users
- Only one service type, best effort, is configured.
- The multiple transmit channel mode is disabled.
- There is no node/link failure.
- There is no piggyback request.
- The protocol proposed will be based on the DOCSIS protocol.

4.3 Modeling of Cable Networks

4.3.1 Introduction

An analytic model was proposed and utilized to study cable network performance with strategic subscribers. In addition to the performance analysis, the model will be used for controller synthesis. The cable network studied was treated as a multi-agent system. The micro-state space was defined as a vector of allocated bandwidth for cable modems, and the macro-state space is defined as a group of social welfare functions of utilities of network participants. In this research, network participants were modeled as utility maximizers and the social agent was modeled as a social welfare maximizer. The quality of service of cable networks was measured with the macro-states.

4.3.2 Multi-stage Model of Cable Networks

The multi-stage model of a cable network system is a 9-tuple: $CN = \{CM_{cn}, T, I, P, A, S, U, SW_{cn}, sw_{cn}\}$, where:

CM_{cn} is a set of *CMs*. $CM_{cn} = \{CM_1, CM_2, \ldots, CM_n\}$, where n is the number of CMs in the system.

T is a set of stages. $T = \{t_1, t_2, \ldots\}$. Under a flat rate, stages are independent. Thus, the multi-stage model can be reduced to a single-stage model by dropping stage index. For analysis of a cable network without the proposed automatic control, the single-stage model is used.

I is a set of interaction rules. For the scope of this paper, there is only one interaction rule. I = {upstream transmission contention resolution}.

P is a set of policies for the cable network. A policy typically is a vector. In this research, the relevant policy is a vector of constant flat service charge rate each subscriber pays, initial back-off windows, and maximum back-off windows.

$A = (A_1 \text{ x } A_2 \text{ x } \ldots \text{ x } A_n)$ is an action space, where A_i is a set of actions (strategies) of CM_i, $i = 1, 2, \ldots, n$. The vector $(a_1, a_2, \ldots, a_n) \in A$ is called a strategy profile. To obey the interaction rule or to disobey are examples of actions for a CM.

$S = (S_1 \text{ x } S_2 \text{ x } \dots \text{ x } S_n)$ is a state space, where $S_i = [0, 1]$ is a set of normalized states of CM_i, s_i: $A_i \text{ x } A_{-i} \text{ x } I \text{ x } P \text{ x } T \rightarrow S_i$, $A_{-i} = (A_1 \text{ x } A_2 \text{ x } \dots \text{ x } A_{i-1} \text{ x } A_{i+1} \text{ x } \dots \text{ x } A_n)$. S_i is used to characterize system microscopic behaviors.

U is a set of utility functions of CMs, where $U = \{u_1, u_2, \dots, u_n\}$ and u_i: $S_i \rightarrow \Re$, $u_i \geq 0$, $u_i(0) = 0$, u_i is increasing and concave.

SW_{cn} is a set of nonnegative social welfare (SW) measurement functions of the cable network. $SW_{cn} = \{sw_1, sw_2, \dots, sw_L\}$. Each element of SW_{cn} is a real-valued social welfare function of the cable network and produces an aggregate of the utility vector $\{u_1, u_2, \dots, u_n\}$. Social welfare functions are used to represent network macroscopic states.

sw_{cn} is a real-valued aggregated social welfare, forming an aggregate of the social welfare function measurement vector $\{sw_1, sw_2, \dots, sw_L\}$. It is analogous to a multi-objective function in the decision-making literature. The function sw_{cn} is a non-decreasing function of sw_j, where $j = 1, 2, \dots, L$, when other social welfare measurement variables are fixed. An example of sw_{cn} is a weighted linear sum of the individual social welfare functions, $sw_{cn} = c_1{*}sw_1 + c_2{*}sw_2 + \dots + c_L{*}sw_L$, where the weight $c_j \geq 0$.

4.3.3 State Space of Cable Networks

The state variables for CMs are called microscopic state variables, and the state variables for the cable network are called macroscopic state variables. The social welfare functions, as macroscopic variables, are used to measure the efficiency and fairness of the cable networks.

Definition 1: The microstate of CM_i, $s_i(a_i, a_{-i}, r, p) \in S_i$, where $a_i \in A_i$, $a_{-i} \in A_{-i}$, $r \in I$ and $p \in P$, is defined as CM_i's percentage share of the bandwidth at the current stage. Because there is externality in cable networks, the state of CM_i depends on not only the actions of CM_i but also on those of the other CMs. A state vector of CM states, *Alloc* = {s_1, s_2, …, s_n}, is called an allocation.

Definition 2: Macro observable states are defined as social welfare functions. The two social welfare functions used in this model are the utilitarian and egalitarian social welfare functions determined by equation (1) and equation (2) respectively based on [7]. The utilization in the field of computer networks is a special case of utilitarian social welfare.

$$sw_{util} = \sum_{u_i \in U,\, s_i \in Alloc} u_i(s_i) \tag{1}$$

$$sw_{egal} = min_{u_i \in U,\, s_i \in Alloc}\, u_i\,(s_i) \tag{2}$$

It is hypothesized that the overall traffic dynamics of the cable network can be characterized by the same four phases identified in [13]. Figure 7 shows the cable network phase transition hypothesized. The identification of four-phases dynamics will be shown in Chapter 5.

Definition 3: The hidden network state space is defined as *PHASE* x *FAIRNESS*, where *PHASE* = {Free Way, Saturated, Bi-State, Congestion} and *FAIRNESS* = {Fair Traffic, Unfair Traffic}. Fair traffic means that the bandwidth allocations are fair in the sense of a fairness criterion, such as proportional fairness.

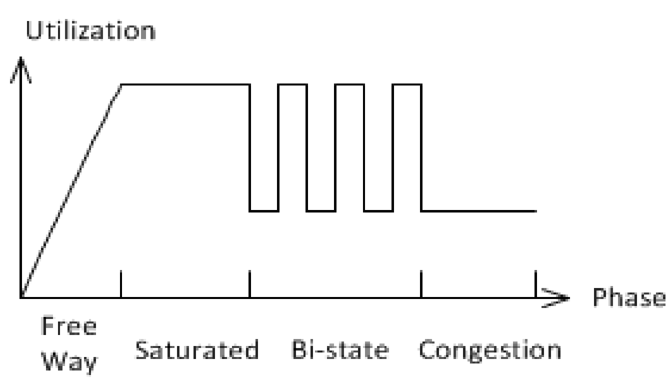

Figure 7. Cable Network Phase Transition Hypothesized

4.3.4 Behavior Models of Agents

This research models a CM as a utility maximizer. Conditional on the network policy p and the actions of the other CMs, CM_i makes local self-interested decisions under an interaction rule based on equation (3).

$$a_{imax} = arg\,max_{a_i \in A_i}\, u_i\,(s_i(a_i, a_{-i}, r, p)) \tag{3}$$

where: $i = 1, 2, \ldots, n$

The social agent of a cable network is modeled as a social-welfare maximizer. Suppose the goal of the cable network is to achieve some QoS to get the desired social welfare objective. For a fixed interaction rule, such as upstream transmission contention resolution, the problem of the cable network can be formulated by equation (4).

$$p_{max} = arg\,max_{p \in P}\, sw_{cn}\,(sw_1(p), sw_2(p), \ldots, sw_L(p)) \tag{4}$$

4.3.5 Process Models

4.3.5.1 Network Process Models

Network process models are Markov models using the hidden states defined by Definition 3, shown in Figure 8 and Figure 9.

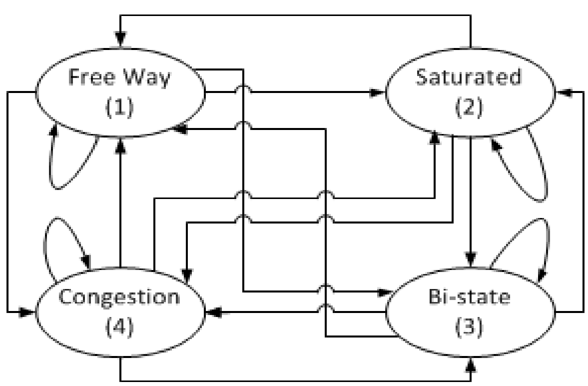

Figure 8. Utilization model

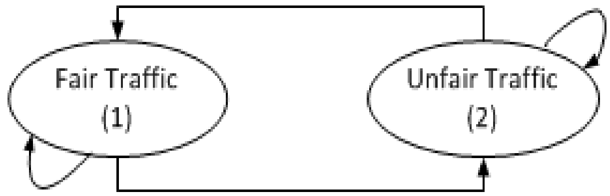

Figure 9. Fairness model

4.3.5.2 CM Process Models

CM process model is presented in the controller synthesis section.

4.4 Observer Design

4.4.1 Introduction

In addition to the observable micro and macro state variables defined in the analytical model, some abstract states were desirable to fulfill the control specification. Through a simulation study, it was identified that the utilization dynamics of cable networks exhibit four phases: free flow, saturated, bi-state, and congestion. The four phases combined with the level of fairness identified were used to construct an abstract state space for the cable network. Therefore, in order to design a controller for a cable network, an observer was needed to evaluate the current network state and traffic flows.

A 9-tuple observer model and a detailed observer design of cable networks were proposed. The observer design was recast into an online traffic classification problem. The designed observer includes four modules: a feature estimator, a flow classifier, a network classifier, and a low-pass filter. The network classifier was designed with a hidden Markov model. Social welfare metrics were used as the feature set of the network classifier and the network classifier quantified the estimated social welfare metrics and used a quantized value as an observation to determine the current network state.

4.4.2 Observer Model

The cable network observer to be constructed is a 9-tuple: C_{obs}={ T_{delta}, *Alloc*, *FL*, *F*, *NF*, *FC*, *NS*, *f*, *g*}, where:

T_{delta} is a set of small identical time intervals. $T_{delta} = \{\Delta_1, \Delta_2, \ldots, \Delta_j, \ldots\}$.

Alloc is a set of bandwidth allocations of the cable network. $\{s_1, s_2, \ldots, s_n\} \in Alloc$, where s_i is the state of CM_i, i.e. CM_i's percentage share of the bandwidth and $i = 1, 2, \ldots, n$, where n is the number of CMs in this cable network.

FL is a set of traffic flows of a cable network.

$F = F_1 \times F_2 \times \ldots F_i \ldots \times F_n$. F_i is the feature space for flow i.

NF is a set of network features, i.e. network feature space.

FC is a set of flow classes. *FC* = {Light, Normal, Heavy} in this dissertation.

NS is the abstract state space of the cable network. In this dissertation, $NS = (PHASE \times FAIRNESS)$.

f is a flow classification assignment, i.e. a map $f: FL \rightarrow FC$.

g is a network state assignment, i.e. a map $g: NF \rightarrow NS$.

4.4.3 Flows & Features

Definition 4: A flow is defined as a sequence of frames which originated from the same CM. A flow is identified with the identifier (ID) of the corresponding CM. This work used positive integers as the IDs of CMs. A flow i is represented by fl_i. $FL = \{fl_1, fl_2, \ldots, fl_n\}$

Definition 5: The feature space of a flow, fl_i, is defined as the state of CM_i, s_i, $\in [0, 1]$, $i = 1, 2, \ldots, n$. Formally, $F_i = \{s_i\}$.

Definition 6: The feature space of a cable network is defined as $NF = \{sw_1, sw_2, \ldots, sw_L\}$. An example of feature space of a cable network is {Utilitarian, Egalitarian, Utilization}, where L =3.

4.4.4 Online-observer Design Specification

The proposed observer for the cable network control system contains four modules: a feature estimator, a flow classifier, a network classifier, and a low-pass filter, shown in Figure 10. The network classifier, which includes a SW quantizer and a network state estimator shown in Figure 11, was designed with a hidden Markov model. The inputs and outputs are the following:

- Observer input: cable network traffic.
- Observer output: the classified flows, the hidden state of the network, and utilization trajectory.
- Feature estimator output: estimated allocation, i.e. a vector of micro states, and estimated social welfare functions
- Flow classifier input: estimated allocation
- Network classifier input: estimated social welfare functions
- Low-pass filter input: estimated social welfare functions

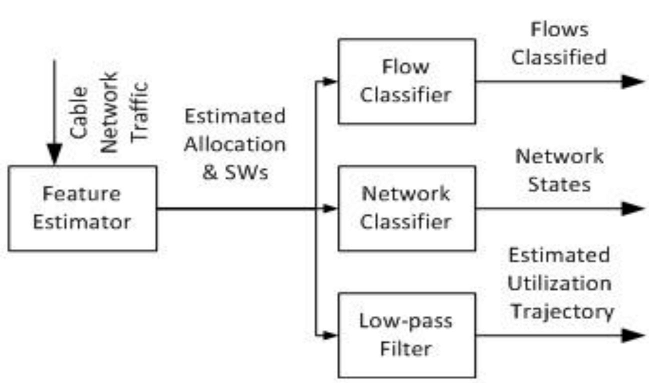

Figure 10. Classifier architecture of a cable network

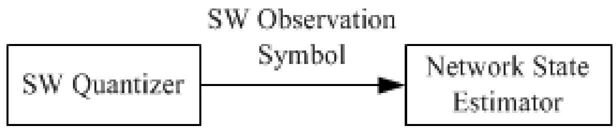

Figure 11. Network classifier components

4.4.5 Online-Observer Component Design

4.4.5.1 Feature Estimation Design

The feature estimator module takes the cable network traffic as inputs, and outputs the estimated allocation of each CM as well as the estimated social welfare metric for the network. The allocation estimation is then used by the flow classifier to classify a flow as light, normal, or heavy. The social welfare metrics is used by the network classifier to determine the traffic phase (freeway, saturated, bi-state, and congestion) and the network fairness (fair, unfair).

4.4.5.1.1 CM Allocation Estimation

As stated above, the feature of a flow, fl_i, is the state of CM_i, $s_i \in [0, 1]$, $i = 1, 2, \dots, n$. Let *cap* be the capacity, in bits per unit of time, of the cable network. The estimation scheme of s_i, $i = 1, 2, \dots, n$, is constructed as follows. Assume that flows corresponding to CMs are independent and the bit count of each CM follows a non-homogenous Poisson process with intensity, μ_i, $i = 1, 2, \dots, n$, which is piece-wise constant through small time intervals.

Let the bit count of the fl_i in Δ_j be $\Delta b_i^{(j)}$, an observation obtained by a flow measurement method in [19]. Similar to the estimate in [20], the intensity (i.e. rate parameter) estimate is determined by equation (5):

$$\hat{\mu}_i^{(j)} = \Delta b_i^{(j)} \Big/ \Delta_j \tag{5}$$

The estimate of s_i, the percentage of bandwidth allocated to CM_i, (represented with $\hat{s}_i$) and the estimate of the network allocation (represented with $\hat{a}$) in Δ_j can be obtained by equation (6) and (7).

$$\hat{s}_i^{(j)} = \hat{\mu}_i^{(j)}/cap \tag{6}$$

$$\hat{a}_i^{(j)} = (\hat{s}_1^{(j)}, \hat{s}_2^{(j)}, \dots, \hat{s}_n^{(j)}) \tag{7}$$

4.4.5.1.2 SW Estimation

The features of a cable network are represented by the social welfare functions based on [7]. The estimated SWs will be fed into the network classifier. The three social welfare functions, utilitarian, egalitarian, and utilization, are estimated as shown in equation (8), (9) and (10). Note that utilization is a special case of the utilitarian social welfare, which will serve as an important feature for the proposed network traffic classification.

$$\widehat{sw}_{util} = \sum_{u_i \in U, i=1}^{n} u_i(\hat{s}_i) \tag{8}$$

$$\widehat{sw}_{egal} = min_{u_i \in U, i \in \{1,2,\dots,n\}} u_i\,(\hat{s}_i) \tag{9}$$

$$\hat{s} = \sum_{i=1}^{n} \hat{s}_i \tag{10}$$

In these equations, $\hat{s}_i$ is the estimate of s_i, the percentage of bandwidth allocated to CM_i i = 1, 2, …n, where n is the number of CMs in the cable network. U is a set of utility functions of CMs. U = $\{u_1, u_2, \dots, u_n\}$ and u_i: $S_i \rightarrow \Re$, $u_i \geq 0$, $u_i(0) = 0$, u_i is increasing and concave. $S_i = \{s_i\}$.

4.4.5.2 Flow Classifier Design

A flow classifier is a map f: $FL \rightarrow FC$. The map requires features of the flow and the flow feature estimates will serve for this purpose. In other words, the flow classifier will use the CM allocation estimation determined by the feature estimator module, and compare the allocation against an upper and lower threshold in order to classify each flow as light, normal, or heavy. The flow classifier uses three parameters:

- s_{pre}: The prescribed state of a CM. In this research, this is defined as the percentage of bandwidth pre-determined by the ISP to be allocated to a CM.
- th_l: Lower allocation threshold factor. $0 < th_l < 1$.
- th_h: Upper allocation threshold factor. $th_l < th_h < 1$.

The classification rules of the flow classifier are the following:

If $s_i < th_l * s_{pre}$, the flow fl_i will be tagged as a light flow.

If $s_i > th_h * s_{pre}$, the flow fl_i will be tagged as a heavy flow.

Otherwise, the flow fl_i will be tagged as a normal flow.

4.4.5.3 Network Classifier Design

A network classifier is a map g: $NF \rightarrow NS$. The network classifier designed in this research takes the estimation of the social welfare functions as input, and outputs the network states. A network state is defined in this research as a traffic utilization state shown in Figure 8 combined with a network fairness state shown in Figure 9. The model includes two parallel sub-models, a utilization HMM and a fairness HMM.

We denote the model parameters of the first sub-model with λ_1 and those of the second with λ_2. $\lambda = (\lambda_1, \lambda_2)$. Model parameters refer to transition probability matrix, output probabilities conditional on a state, and initial state probabilities. Initially, these probabilities are guessed by the model implementers, and then they can be adapted over time.

The two sub-models in Figures 8 and 9 are combined to get another HMM, shown in Figure 12, so that techniques for HMMs can be readily applicable. The HMM obtained is an 8-node complete directed graph plus self-loops at each node. Each HMM state is identified by a pair (i_1, i_2). The number i_1 corresponds to the utilization HMM and the number i_2 corresponds to the fairness HMM.

To apply the HMM techniques presented in [29], the states in Figure 12 were relabeled with values {1, 2, …, N}, N = 8. The state ID variable i was calculated as $i = 2*i_1 - (2 - i_2)$, i_1= 1, 2, 3, 4, i_2 = 1, 2. For example, state (2, 1) which corresponds to (Saturated, Fair) was relabeled as 3. It was assumed that the underlying Markov chain is homogeneous. HMM is a triple $\lambda = (A_{hmm}, B_{hmm}, \pi)$. A_{hmm}, B_{hmm}, and π can be determined by λ_1 and λ_2. $A_{hmm} = \{a_{ij}\}$ is the state transition probability matrix, $B_{hmm} = \{b_j(k)\}$ is the observation symbol probability distribution conditional on the state j, $k = 1, 2, \ldots, M$, and $\pi = \{\pi_i\}$ is the initial state distribution [29].

4.4.5.3.1 SW Quantizer

In order to use the techniques of HMM and simplify the problems, especially the implementation problems, the vector of the social welfare functions were quantized into a set of discrete values. In other words, the social welfare vector space was mapped to a 1-dimension discrete space. Another advantage of quantization is for interpretation purpose. For example, the ratio of utilitarian to egalitarian can be used to measure fairness. Let this 1-dimension discrete space be {v_1, v_2, …, v_M}. {v_1, v_2, …, v_M} will serve as the observation symbols emitted by the HMM.

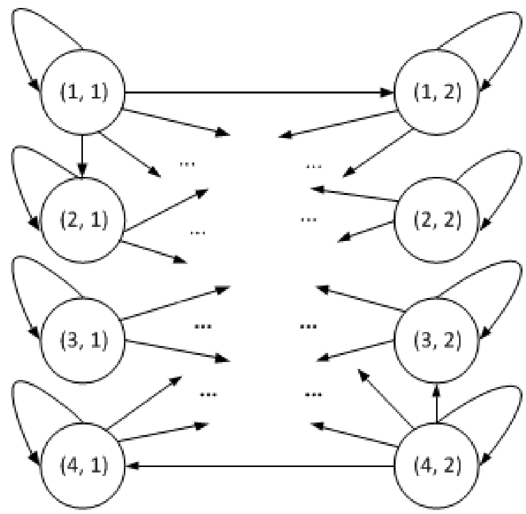

Figure 12. Network state diagram, a HMM

4.4.5.3.2 *Network State Estimator*

Network state estimator is used to recognize dynamic phase changes of cable networks automatically. It has the following properties:

- Dynamical.
- Recursive.
- Adaptive & learning-based.

4.4.5.4 Low-pass Filter Design

A low-pass filter is used to smooth the utilization chart in the free-flow or saturated phase and the resulting curve is used for controller synthesis. The low-pass filter is designed with a EWMA defined as

$$\begin{cases} \tilde{s}_t = (1-\theta)s_t + \theta\tilde{s}_{t-1} \\ \tilde{s}_0 = s_0 \end{cases} \quad (11)$$

where s_t is the observed network utilization, i.e. the sum of the bandwidth allocation ($s_t = \sum_{i=1}^{n} y_{it} / cap$) at time t and $\tilde{s}_t$ is the estimated utilization at time t. $0 \leq \theta \leq 1$. Smaller θ emphasizes recent observations and larger θ makes the process smoother. In this dissertation, θ is selected to be 7/8 as in [41].

4.5 Controller Synthesis

4.5.1 Introduction

Controller synthesis is done through the methods of mechanism design and control theory. Mechanism design means to devise a mechanism in order to induce a new game, which results in desired outcomes. There are three possible design subspaces: interaction rule space, allocation rule, and policy (including payment) space. Modification of any of the variables in these subspaces could induce a new game. In this dissertation, the interaction rule will not be touched, that is, the reservation Aloha will not be changed. The allocation rule will be kept the same as before, i.e. CMTS allocates bandwidth to whatever user whose request has been successfully received. Mechanism design will be performed by means of heuristic design methodology to trade off optimal efficiency for simple design and implementation.

Mechanism design can be one-shot or dynamic. Because of system uncertainties, such as model uncertainties (including un-modeled dynamics and parameter uncertainties) and disturbances, one-shot design is virtually impractical. Dynamic mechanism design allows the system to enter the designed equilibriums in finite time steps instead of one-shot. Thus, control theory methodologies can play a very important role. In this dissertation, mechanism design of cable networks is treated from the perspective of control theory.

There exist several senses of fairness, such as proportionally fair and max-min fair. This dissertation only considers one sense of fairness, proportional fairness. It is assumed that the system is proportionally fair if everyone obtains his/her intended bandwidth as mentioned in [57]. Then, the desired goal of the system becomes maximization of network utilization under the constraint of proportional fairness. Proportional fairness allows us not to deal with fairness separately because the cable network is fair if the cable network is in Free Way or Saturated state and the cable network is unfair if the cable network is in Bi-state or Congestion state. Removing Bi-state or Congestion implies removing unfairness.

The proposed controllers of cable networks have the following properties:

- Sub-optimal network utilization
- Agent-based distributed control
- Sliding mode control, a special case of adaptive control and robust control
- State feedback control
- Hybrid control

4.5.2 Controller Design Specifications

The following are the specifications for controller design:

- Control signal (Controller output): Rate of charge for bandwidth
- Controller input: The classified flows, the hidden state output, and network utilization by the observer.
- The reference specifications (goals): sub-optimal network utilization.
- Control communication specifications: Control signal is carried within the control packet or MAP.

4.5.3 Behavior Models Revisited

By introducing dynamic pricing, an agent's utility function is modeled as a quasi-linear function of the following form:

$$u_i(y_i/cap) = v_i(y_i) - p * y_i \tag{12}$$

where, y_i (= s_i * cap) is introduced for easier interpretation. v_i: $\Re^+ \cup \{0\} \rightarrow \Re^+ \cup \{0\}$, $v_i(0) = 0$, v_i is increasing and concave. $i = 1, 2, \ldots, n$, y_i is the bandwidth obtained by this agent, v_i can be interpreted as a valuation function of this agent.

The objective of the social agency is to maximize utilization s under the constraints of proportional fairness, that is

$$\begin{cases} \max_{s \in \{s_1, s_2, \ldots, s_n\}} s \\ subject\ to \\ \quad proportional\ fairness \end{cases} \tag{13}$$

4.5.4 Characteristics of Market Bandwidth Demand Curve

Market demand for a particular good is the sum of all the individual demands for the same good [82]. Dupuit's laws of demand state that demand curves slope downward and are of convex shape because of diminishing marginal utilities [83]. Thus, a demand curve reflects the prices the customers are willing to pay for the specific levels of the goods demanded. A downward demand curve for a cable network market means that a cable network user is willing to pay more for light use than for heavy use. That is, a user values light use more than heavy use because of the property of his/her diminishing marginal utility. Based on bandwidth quantity demand, the demand curve is divided into three regions as shown in Figure 13. Each of regions corresponds to light use, normal use, and heavy use. Let the maximal bandwidth demand of the regions be s_l, s_n, and s_h respectively and the corresponding prices be p_l, p_n, and p_h respectively.

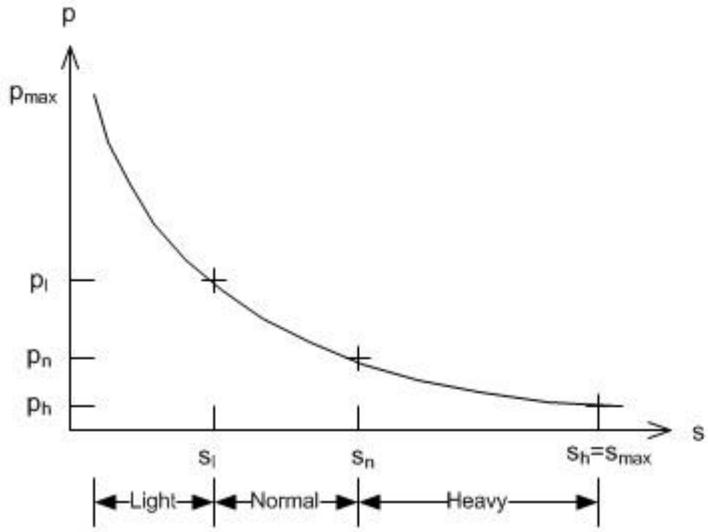

Figure 13: Bandwidth demand curve

4.5.5 Controller Design of Cable Networks

4.5.5.1 Cable Network Controllers

Mechanism design is used to synthesize the controller. As mentioned earlier, dynamics of price implies dynamics of mechanism in this dissertation. The dynamics of cable networks can be also modeled as the following:

$$ds/dt = f_\sigma(y, p) \tag{14}$$

where $f_\sigma(y, p)$ is the dynamics of the cable network. A switched-system model is used. Discrete network states are mapped to two switched subsystems, i.e. $\sigma = 1, 2$.

Switched subsystem 2 ($\sigma = 2$): Cable network is in the Bi-State or Congestion state.

This controller achieves **congestion exclusion** of cable networks. The control signal (i.e. the price) is constructed as

$$p_k = p(t_k) = K_2 \tag{15}$$

where $K_2 \geq p_{max} = max_s D^{-1}(s)$. K_2 is the controller gain for the switched subsystem 2 and it can be thought of as a congestion fee.

Switched subsystem 1 (σ = 1): Cable network is in the Free Way or Saturated state.

This controller achieves **congestion avoidance** of cable networks. It is assumed that the social agency does not know the valuation functions of agents, but has some idea of the demand curve for this cable network.

Let s_{max} be the prescribed or provisioned network utilization when all CMs are homogenous. Assume that s_{max} is known. To prevent congestion, choose $s_{th} < s_{max}$ to be a desired utilization, where s_{th} is selected close to s_{max}. Define $s_d = s - s_{th}$. Let $\varepsilon(s_d)$ be a Heaviside unit step function: $\varepsilon(s_d) = 1$ if $s_d > 0$, $\varepsilon(s_d) = 0$ otherwise. Here, $s - s_{th} = 0$ is called a sliding surface or sub-efficiency frontier. Then, based on sliding mode control theory, the control signal (i.e. the price) is constructed as

$$p_k = p(t_k) = K_1 * \varepsilon(s_d) \tag{16}$$

where $K_1 \geq p_n$. K_1 is the controller gain for the switched subsystem 1.

Control law (16) is intended to make the cable network sub-efficient and trades off efficiency for simplicity. It is needed because of four reasons: (1) Finding the Pareto frontier is NP hard. (2) Because of system uncertainties, determination of payment rules to keep the system on the Pareto frontier will be very complicated. (3) The system may experience frequent large utilization swings between Subsystem 1 and Subsystem 2. (4) Light users need to be well protected.

4.5.5.2 CM Process Model

As stated before, each agent is a utility maximizer. Thus, the dynamics of each agent can be modeled by following gradient play, that is

$$x_{i,k} = x_{i,k-1} + \tau_i * \kappa_{i,k}(y_{i,k-1}) * (\nabla v_i(y_{i,k-1}) - p_k) \quad (17)$$

where $x_{i,k}$ is the transmission rate (bits/s) determined by agent i at time k, τ_i is the time step, which is equal to or greater than the minimum of MAP time, $\kappa_{i,k} > 0$ is a parameter, which measures the sensitivity of the transmission rate change with time to the utility change with the allocation. Here, $\kappa_{i,k}$ is private to agent i and could be a constant. This equation can be considered as a CM process model which determines the strategy of CMs. It is consistent with the fact that any user will transmit frames as fast as possible to maximize his/her utility and the network will blow up when the dynamic price p_k is zero.

4.5.6 Charging Rule

It is assumed that a flat rate is paid by each subscriber as usual. If $p_k \geq 0$ and CM_i is a heavy user, charge CM_i an additional fee; otherwise, there is no charge for everyone.

4.5.7 Protocol Modification

The modification of the control frame, MAP, is small in this dissertation. The only modification is to add a field, the bandwidth unit price of dollars per the unit of Kbps (Kilobits per second) within the time interval.

Upstream channel ID	UCD Count	# of element	BW Price	Alloc Start Time	ACK Time
Ranging Backoff Start	Ranging Backoff End	Data Backoff Start	Data Backoff End	MAP Info Elements	

Figure 14. Modified MAP format

CHAPTER 5

THEORETIC ANALYSIS AND COMPUTER SIMULATION OF CABLE NETWORKS AND THE CABLE NETWORK CONTROL SYSTEM

5.1 Introduction

Theoretical analyses and computer simulations were developed for each phase to validate the design and various other aspects of the research. The model designed in the first phase of this research was used to perform theoretical analysis of cable networks under flat pricing to show the problems of congestion and subsidization.

The proposed control system was evaluated by studying the properties of a cable network control system, such as controllability, stability, incentive compatibility, voluntary participation, and sub-efficiency. In order to demonstrate the effectiveness of the proposed cable network control system, the NS2 network simulator was used to simulate the cable network performance both without and with the proposed control solution.

5.2 Theoretical Analysis

5.2.1 Introduction

Two scenarios were analyzed using the model proposed in this research. In the first scenario, it was assumed that the CMs can choose whether or not to follow the congestion control mechanism present in the DOCSIS specification. The second scenario assumed that all CMs obey the specification, but there are some heavy users in the system. The results of these scenarios were used to validate the conclusion that the current control systems available for cable networks are not adequate to ensure efficiency and fairness in cable networks.

5.2.2 Scenario 1: Cable Networks with Obeying or Disobeying CMs

The purpose of this scenario analysis was to predict the consequences of the potential problem that network participants are able to implement protocols which deviate from standard protocol specifications for personal benefit. For this scenario, assume that the cable network is under heavy load such that all CMs always have data to transmit. In addition, all CMs can either obey or disobey the upstream transmission contention resolution. The allocation of the network is $(s_1, s_2, \ldots, s_n)$ and the utility vector of the network is $(u_1, u_2, \ldots, u_n)$.

Let $u_{i[allobey]}$ represent the utility of CM_i when all CMs obey. Let $u_{i[disobey]}$ represent the utility of CM_i when CM_i disobeys and the remaining CMs obey, where $i = 1, 2, \ldots, n$. It has been proven that $u_{i[allobey]} < u_{i[disobey]}$ in the proof of proposition 1 provided in section 5.2.4. The analysis shows that the Nash equilibriums are the strategy profiles in which at least one CM disobeys. That is, all the strategy profiles are Nash equilibriums except the prescribed strategy profile, in which all CMs obey. There are $2^n - 1$ Nash equilibriums.

Table 1 shows some social welfare measurements for the various classes of action profiles analyzed, with the following action profile classes:

- *AP1*- all CMs obey
- *AP2*- only one CM, CM_i, disobeys
- *AP3*-two or more CMs disobey

In Table 1, USW is the utilitarian social welfare and ESW is the egalitarian social welfare. For the equilibriums in which two or more CMs disobey, all social welfare measurements are zero since the cable network is completely congested. For the equilibriums in which one and only one CM disobeys, the utilitarian social welfare may be desirable but the egalitarian social welfare is zero. In these cases, while some traffic is successfully transmitted, the principle of fairness is severely violated. It can be concluded from the results of this analysis that the goals of the CMs conflict with those of the cable network under the flat-rate policy.

TABLE 1

Social Welfares of the Classes of Action Profiles

Action Profile	**USW**	**ESW**	**NE?**
AP1	$n{*}u_{i[allobey]}$	$u_{i[allobey]}$	No
AP2	$u_{i[disobey]}$	0	Yes
AP3	0	0	Yes

5.2.3 Scenario 2: Cable Networks with Heavy and Light Users

The purpose of this scenario analysis is to theoretically explain how light users subsidize heavy ones and heavy users can cause severe network congestion under a flat pricing scheme, which is reported in a study of commercial/experimentation broadband networks [11] [12]. Assume that heavy users have two strategies to choose from, heavy use or light use, while light users can only choose the light use strategy. The analysis demonstrates that the strategy profile in which all heavy CMs select the heavy use strategy is the Nash equilibrium. Let the probability that heavy CMs are idle be α ($0 < \alpha < 1$) and the probability that light CMs are idle be β ($0 < \beta < 1$). $\alpha < \beta$. Let the number of light CMs be n_1 and the number of heavy CMs be n_2, where $n_1 + n_2 = n$.

Suppose k ($n_2 - 1 \geq k \geq 0$) heavy CMs choose the heavy use strategy. Let another heavy CM's state or share of the bandwidth be $s_{(k+1)H}$ if it chooses the heavy use strategy and $s_{(k+1)L}$ if it chooses the light use strategy. Using properties of the DOCSIS collision resolution scheme, it has been proven that $s_{(k+1)H} > s_{(k+1)L}$ resulting in $u(s_{(k+1)H}) > u(s_{(k+1)L})$ in the proof of proposition 2 provided in section 5.2.4. This result shows that all heavy CMs select the heavy use strategy since selecting the heavy use strategy results in a strictly higher utility value than selecting the light use strategy when the strategies of the other CMs are fixed. Therefore, the strategy profile in which all heavy CMs select the heavy use strategy is the Nash equilibrium.

Let s_{heq} be the state of any heavy CM and s_{leq} be that of any light CM under the Nash equilibrium. Let $allocsum_{eq}$ be the network utilization under the Nash equilibrium. The proof of proposition 3, given in section 5.2.4, shown that that s_{heq} > s_{leq} and that when n_2 grows, $allocsum_{eq}$ approaches to zero. Therefore, it can be concluded that under the Nash equilibrium the utilitarian social welfare could be close to zero when too many heavy CMs exist. It can be concluded from this scenario that even when the interaction rules of the cable network are obeyed, the controls of the cable network may not adequately ensure the fairness of the CMs and an appropriate utilitarian social welfare under the flat-rate policy.

5.2.4 Proofs of Lemmas & Propositions

Figure 15 shows a state diagram of contention resolution, in which r is a non-negative integer drawn from a uniform distribution in the interval $[0, w_k)$, where w_k denotes the most current stage-k back-off window, k = 0, 1, 2, …, m, m+1 is the number of maximum back-off window stages.

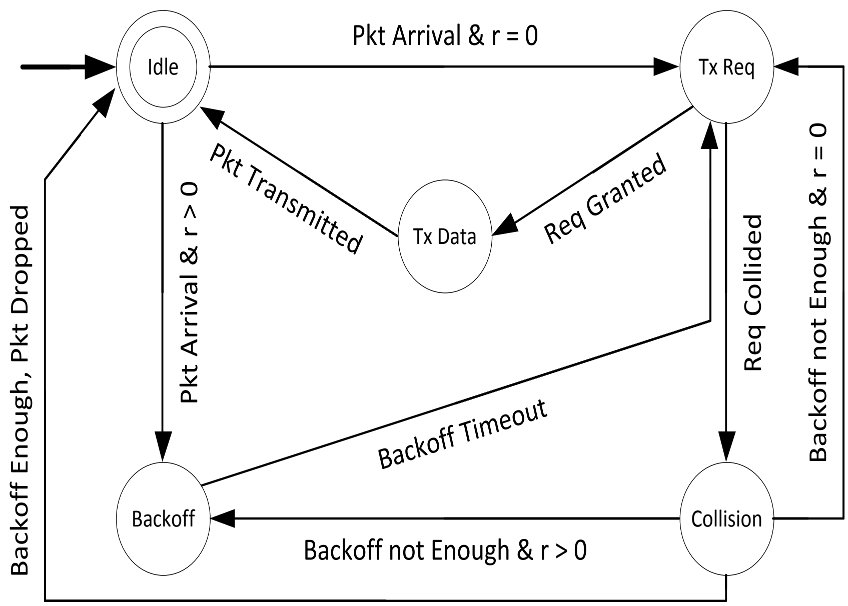

Figure 15. State diagram of contention resolution

Let $0 < p^o_{bj} < 1$ be the back-off probability of CM_j and $0 < p^o_{ti} < 1$ be the request transmission probability of CM_i when all CMs are obedient. The percentage of bandwidth obtained by CM_i is

$$s_{i[allobey]} = p^o_{ti} * \prod_{j=1, j\neq i}^{j=n} p^o_{bj} \tag{18}$$

where $p^o_{bj} = \sum_{k=0}^{k=h} a^o_k \, p_{bj}(w_k)$, $p_{bj}(w_k)$ is the back-off probability of CM_j with the back-off window w_k, $w_{k-1} < w_k$ and $p_{bj}(w_{k-1}) < p_{bj}(w_k)$, w_0 is the initial back-off window size, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$, $j \neq i$, $0 < a^o_k \leq 1$, $h \leq m$, $\sum_{k=0}^{k=h} a^o_k = 1$, a^o_k represents the probability for an obedient CM_j to select the back-off window, w_k.

Let $0 < p^d_{bj} < 1$ be the back-off probability of CM_j, one of the obedient CMs, when CM_i disobeys and the rest of CMs obey. The percentage of bandwidth obtained by CM_i is

$$s_{i[disobey]} = \prod_{j=1, j\neq i}^{j=n} p^d_{bj} \tag{19}$$

where $p^o_{bj} = \sum_{k=0}^{k=m} a^d_k \, p_{bj}(w_k)$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$, $j \neq i$, $0 < a^d_k \leq 1$, $\sum_{k=0}^{k=m} a^o_k = 1$, a^d_k represents the probability for an obedient CM_j to select the back-off window, w_k, when CM_i disobeys and the remaining CMs obey.

If a CM disobeys, the rest of CMs, which always get collisions, get equal chances to enter back-offs with any possible w_k. Therefore, $a^d_k = 1 / (m + 1)$ for any k, $0 \leq k \leq m$.

Based on upstream transmission contention resolution, a smaller back-off window is always selected before a larger one. Thus, $a^o_0 \geq a^o_1 \geq a^o_2 \geq \ldots \geq a^o_h > 0$.

Lemma 1: If $h < m$, or $h = m$ and $\exists k'$ ($0 \leq k' < h$ and $a^o_{k'} > a^o_{k'+1}$), then, $a^o_0 > 1 / (m + 1)$

Proof of Lemma 1: We use proof by exhaustion and proof by contradiction.

Case (1): Let $h < m$ and $a^o_0 \leq 1 / (m + 1)$, then $a^o_0 + a^o_1 + \ldots + a^o_h \leq (h + 1) * a^o_0 \leq (h + 1) / (m + 1) < 1$. This contradicts to $\sum_{k=0}^{k=h} a^o_k = 1$.

Case (2): Let $h = m$, $\exists k'$ ($0 \leq k' < h$ and $a^o_{k'} > a^o_{k'+1}$), and $a^o_0 \leq 1 / (m + 1)$. Then, $a^o_{k'+1} < 1 / (m + 1)$ for some k'. Since $a^o_0 + a^o_1 + \ldots + a^o_h = a^o_0 + \ldots + a^o_{k'} + a^o_{k'+1} + \ldots + a^o_m < (k' + 1) / (m + 1) + (m - k') / (m + 1) = 1$. This contradicts to $\sum_{k=0}^{k=h} a^o_k = 1$.

Proposition 1: $u_{i[allobey]} < u_{i[disobey]}$

Proof of Proposition 1: When all CMs obey, contention resolution process indicates that $h < m$, $h = m$ and $\exists k'$ ($0 \leq k' < h$ and $a^o_{k'} > a^o_{k'+1}$), or $h = m$ and $a^o_0 = \ldots = a^o_k = \ldots a^o_m = 1 / (m + 1)$. Based on Lemma 1, the distribution of a^d_k has first-order stochastic dominance over that of a^o_k. With the theory of stochastic dominance, we have

$$\begin{cases} \prod_{j=1, j \neq i}^{j=n} p^o_{bj} \leq \prod_{j=1, j \neq i}^{j=n} p^d_{bj} \\ s_{i[allobey]} = p^o_{ti} \prod_{j=1, j \neq i}^{j=n} p^o_{bj} < 1 * \prod_{j=1, j \neq i}^{j=n} p^d_{bj} = s_{i[disobey]} \end{cases} \tag{20}$$

Thus, $u_{i[allobey]} < u_{i[disobey]}$ according to the properties of the utility functions stated in the model.

Let $p_{b\alpha\alpha}$ be the back-off probability of a heavy user rather than k+1 th user when k+1 th user select heavy use strategy. Let $p_{b\beta\alpha}$ be the back-off probability of a light user when k+1 th user select heavy use strategy. Let $p_{b\alpha\beta}$ be the back-off probability of a heavy user when k+1 th user select light use strategy. Let $p_{b\beta\beta}$ be the back-off probability of a light user rather than k+1 th user when k+1 th user select light use strategy. Let $p_{t\alpha}{}^{k+1}$, $p_{t\beta}{}^{k+1}$ be the request transmission probabilities of k+1 th user when this user selects heavy use strategy and light use strategy respectively.

$$\begin{cases} s_{(k+1)H} = (1-\alpha)p_{t\alpha}^{k+1} * \left((1-\alpha)p_{b\alpha\alpha} + \alpha\right)^{k} * \left((1-\beta)p_{b\beta\alpha} + \beta\right)^{n-k-1} \\ = (1-\alpha)p_{t\alpha}^{k+1} * (1 - p_{c\alpha}^{k+1}) \\ s_{(k+1)L} = (1-\beta)p_{t\beta}^{k+1} * \left((1-\alpha)p_{b\alpha\beta} + \alpha\right)^{k} * \left((1-\beta)p_{b\beta\beta} + \beta\right)^{n-k-1} \\ = (1-\beta)p_{t\beta}^{k+1} * \left(1 - p_{c\beta}^{k+1}\right) \end{cases} \quad (21)$$

where $p_{c\alpha}{}^{k+1}$ and $p_{c\beta}{}^{k+1}$ are k+1 th user's collision probabilities when this user selects heavy use and light use strategies respectively.

Proposition 2: $s_{(k+1)H} > s_{(k+1)L}$

Proof of Proposition 2: If a user switches from light use strategy to heavy use strategy and the other users' loads do not change, then the other users have the same or higher collision probabilities. It is implied that the other users have the same or higher back-off probability. Therefore, we have $p_{b\alpha\alpha} \geq p_{b\alpha\beta}$ and $p_{b\beta\alpha} \geq p_{b\beta\beta}$. Thus, $(1 - p_{c\alpha}{}^{k+1}) \geq (1- p_{c\beta}{}^{k+1})$, which implies $p_{c\alpha}{}^{k+1} \leq p_{c\beta}{}^{k+1}$.

Back-off process indicates that the request transmission probability decreases or keeps equal with the collision probability increasing. Thus, $p_{t\alpha}{}^{k+1} \geq p_{t\beta}{}^{k+1}$.

Therefore, $s_{(k+1)H} > s_{(k+1)L}$ since $\beta > \alpha$.

Let $p_{c\alpha}$ be the collision probability of a heavy user and $p_{c\beta}$ be collision probability of a light user. Let $p_{t\alpha}$ be the request transmission probability of a heavy user and $p_{t\beta}$ be the request transmission probability of a light user. $p_{b\alpha}$ and $p_{b\beta}$ are the corresponding back-off probability. $p_{t\alpha} + p_{b\alpha} = 1$ and $p_{t\beta} + p_{b\beta} = 1$. Let $c = ((1-\alpha)p_{b\alpha} + \alpha)^{n_2-1} * ((1-\alpha)p_{b\beta} + \beta)^{n_1-1}$, then:

$$\begin{cases} s_{heq} = c * (1-\alpha)p_{t\alpha} * ((1-\alpha)p_{b\beta} + \beta) \\ = c * (1-\alpha)p_{t\alpha} * p_{b\beta} + c * p_{t\alpha} * \beta(1-p_{b\beta}) - c * \alpha * p_{t\alpha} * \beta 1 - p_{b\beta} \\ = c * (1-\alpha)(1-p_{b\alpha}) * p_{b\beta} + c * (1-p_{b\alpha}) * \beta\left(1-p_{b\beta}\right) - c * \alpha * (1-p_{b\alpha}) * \beta\left(1-p_{b\beta}\right) \\ s_{leq} = c * (1-\beta)p_{t\beta} * ((1-\alpha)p_{b\alpha} + \alpha) \\ = c * (1-\beta)p_{t\beta} * p_{b\alpha} + c * p_{t\beta} * \alpha(1-p_{b\alpha}) - c * \beta * p_{t\beta} * \alpha(1-p_{b\alpha}) \\ = c * (1-\beta)\left(1-p_{b\beta}\right) * p_{b\alpha} + c * \left(1-p_{b\beta}\right) * \alpha(1-p_{b\alpha}) - c * \beta * \left(1-p_{b\beta}\right) * \alpha(1-p_{b\alpha}) \end{cases}$$

Proposition 3: $s_{heq} > s_{leq}$

Proof of Proposition 3:

$$\begin{aligned} & s_{heq} - s_{leq} = c * (1-\alpha)(1-p_{b\alpha}) * p_{b\beta} + c * (1-p_{b\alpha}) * \beta(1-p_{b\beta}) \\ & -(c * (1-\beta)(1-p_{b\beta})p_{b\alpha} + c * (1-p_{b\beta}) * \alpha(1-p_{b\alpha})) \\ & = c * ((p_{b\beta} - p_{b\alpha}) - (\alpha p_{b\beta} - \beta p_{b\alpha}) + \alpha p_{b\alpha}p_{b\beta} - \beta p_{b\alpha}p_{b\beta}) \\ & +c * ((\beta-\alpha) - \left(\beta p_{b\beta} - \alpha p_{b\alpha}\right) - \left(\beta p_{b\alpha} - \alpha p_{b\beta}\right) + (\beta p_{b\alpha}p_{b\beta} - \alpha p_{b\alpha}p_{b\beta})) \\ & = c * ((\beta - \beta p_{b\beta} + p_{b\beta}) - (\alpha - \alpha p_{b\alpha} + p_{b\alpha})) \\ & = c * (\beta + p_{b\beta}(1-\beta)) - (\alpha + p_{b\alpha}(1-\alpha))) \end{aligned}$$

Let $p_{b\beta} = p_{b\alpha} + \Delta$. Then

$$\begin{cases} s_{heq} - s_{leq} = c * ((\beta + (p_{b\alpha} + \Delta)(1-\beta)) - (\alpha + (p_{b\alpha}(1-\alpha)) \\ = c * ((\beta-\alpha)p_{t\alpha} + \Delta(1-\beta)) \end{cases} \tag{22}$$

Given the number of collisions (including no collision), the ratio of the collided requests to the total requests of a light user is the same as or higher than that of a heavy user. A CM's back-off window size increases if and only if a collision is seen by this CM. Using the similar arguments to those in Proposition 2, $\Delta \geq 0$. Therefore, from equation (9), $s_{heq} > s_{leq}$. since $c > 0$, $\beta > \alpha$, $\beta \leq 1$, and $p_{t\alpha} > 0$.

Proposition 4: The network utilization is low when the number of heavy users is too large.

Proof of Proposition 4: The network utilization can be expressed as:

$$\begin{cases} allocsum_{eq} = \sum_{k=1}^{n} s_k \\ = \sum^{n_1} (1-\beta) p_{t\beta} * ((1-\alpha) p_{b\alpha} + \alpha)^{n_2} * ((1-\beta) p_{b\beta} + \beta)^{(n_1-1)} \\ + \sum^{n_2} (1-\alpha)\, p_{t\alpha} * \left((1-\alpha) p_{b\alpha} + \alpha\right)^{n_2-1} * ((1-\beta) p_{b\beta} + \beta)^{n_1} \end{cases}$$

Ignoring the allocation of bandwidth to light users, the network utilization can be approximated with

$$allocsum_{eq} \approx n_2 (1-\alpha) p_{t\alpha} ((1-\alpha) p_{b\alpha} + \alpha)^{(n_2-1)} \quad (23)$$

When n_2 grows, $allocsum_{eq}$ approaches to zero since $(1-\alpha)p_{b\alpha} + \alpha < 1$.

5.3 Simulation Studies

5.3.1 Introduction

Three simulation studies have been performed in this research. The first simulation study was performed to show that there is a four-phase dynamic traffic pattern in cable networks similar to the traffic pattern observed in other networks. The second simulation was designed to confirm the theoretical results and conclusions regarding the properties of a cable network in its steady state under flat pricing. The last simulation was carried out to demonstrate soundness of observer design and to validate the effectiveness of the proposed cable network control system. The results of the last simulation are presented in the section “5.4 Evaluation of Cable Network Control System”.

5.3.2 Simulation Models & Parameter Configuration

In this research, a set of NS2 DOCSIS models was created to simulate the dynamic traffic pattern and characteristics of cable networks. NS2 version 2.35 was used [84]. The DOCSIS models inherited the existing NS2 MAC models.

As in the theoretic analysis, it was assumed that CMTS only provides best-effort service. The MAP time, which is the number of time slots (including both data mini-slots and request mini-slots) described in a single MAP, was set to a constant. The data mini-slots and request mini-slots were also constant. In order to be consistent with the theoretical study, it was also assumed that there was only one request per mini-slot, which is large enough to hold a request frame, and that the high layer packets ready to transmit exactly fit into the data slots in each single MAP. It was assumed that the simulated CMTS sends a MAP at the beginning of each MAP time, similar to the assumptions made in [8].

The parameters of the simulated cable network are shown in Table 2. The traffic model and its parameters are shown in Table 3.

TABLE 2

Parameters of the Simulated Cable Network

Name	Value
Cable network upstream bandwidth	10Mbps
Link delay	1ms
Minimum backoff window	5
Maximum backoff window	5
Upstream mini-slot size	64 bytes
# of upstream data mini slots	16
Maximum retry	1

TABLE 3

Traffic Model and Parameters for the Simulated Cable Network

Name	Value
Traffic model	Constant bit rate (CBR)
Bit rate of heavy users	1Mbps
Data packet size	1000 Bytes

Figure 14 shows a typical packet/frame sequence diagram of DOCSIS for CMs. It describes the timing scheme of the simulation.

5.3.3 Cable Network Traffic Pattern Dynamics

The purpose of this scenario is to investigate whether a dynamic four-phase traffic pattern (free way, saturated, bi-state, congestion) is present in cable networks. In this scenario, it was assumed that all users are heavy users. To be able to observe the network dynamics, selecting time scale was crucial. A time scale of 0.15 seconds was chosen for this scenario. The NS2 simulation shows that a dynamic four-phase network traffic pattern observed in other networks also exists in cable networks, as shown in Figure 17.

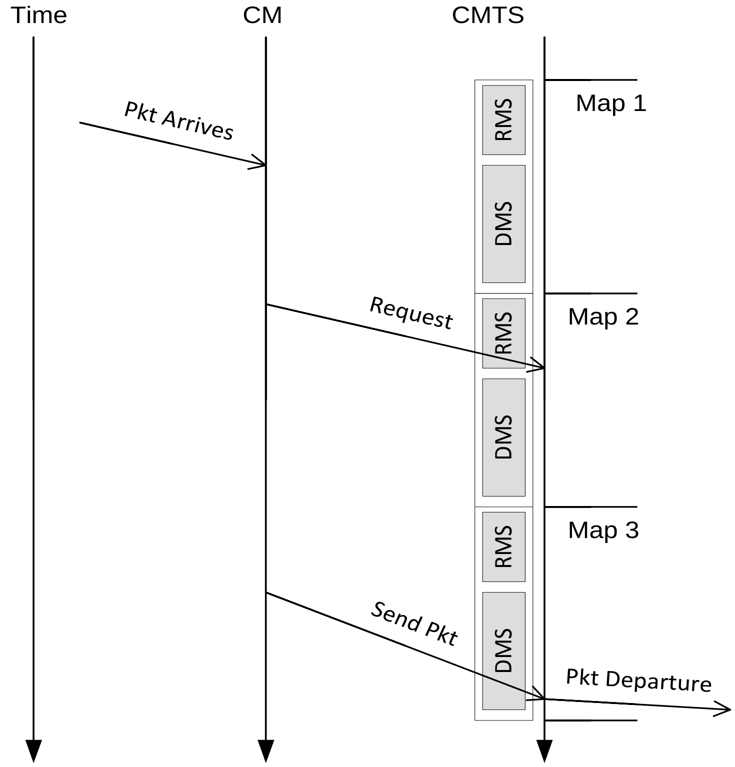

Figure 16. A frame sequence diagram of DOCSIS

Figure 17. Four-phase cable network traffic dynamics

From time 0 to 7.5 seconds, the cable network was in the freeway phase. From time 7.5 to 10 seconds, the cable network was in saturated state. From 10 to 18 seconds, the cable network was in bi-state phase, in which the cable network transitions from high utilization to a low utilization frequently and discontinuously. Bi-state is an unstable state, which is caused by sudden increase and decrease of traffic. From 18 to 30 seconds, the cable network was in congestion phase.

5.3.4 Fairness Problem of Cable Network*s*

The purpose of this scenario is to investigate the difference between the bandwidth allocated for heavy users and that for light users in order to evaluate how heavy users affect the utilization of cable networks. In this scenario, the total number of users was set to 8, in which 4 were heavy users and 4 were light users. Table 4 shows the comparison between the percentage of bandwidth allocated for a heavy user and that for a light user. It can be seen that a heavy user acquires significantly larger bandwidth allocation. This experiment confirms that, under flat-rate policy, light users subsidize heavy users.

TABLE 4

Bandwidth Allocations between a Heavy User and a Light User

User Type	**Bit Rate (Mbps)**	**Percentage of Bandwidth Allocated (%)**
Heavy	1	1.8
Light	0.1	0.8

5.3.5 Utilization Problem of Cable Networks

In this scenario, all users are heavy users and the number of them varies. The network utilization versus the number of heavy users is shown in Figure 16. It can be observed that there exists a CM size such that the cable network reaches an optimum utilization. In this experiment, the optimal number of CMs is about 15. With this optimum, the cable network is in the saturated phase. Increasing number of CMs beyond the optimal number of CMs results in a decline of cable network utilization.

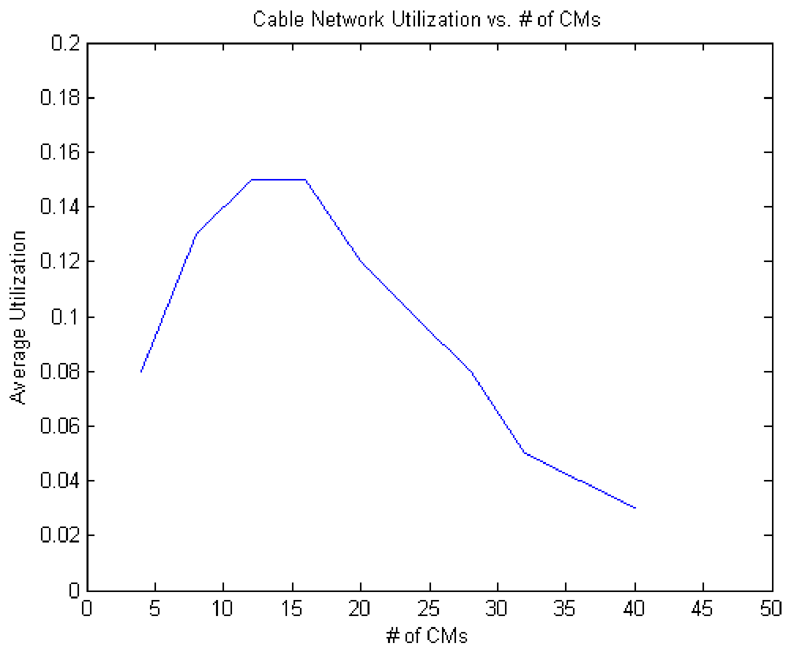

Figure 18: Cable network utilization vs. number of heavy users

5.4 Evaluation of Cable Network Control System

5.4.1 Introduction

The properties of the cable network control system, such as controllability, stability, sub-efficiency, incentive compatibility, and voluntary participation were evaluated. Controllability theory in supervisory control for discrete event systems was applied to show the controllability of cable networks. A Lyapunov function was created to prove the stability of the proposed control of cable networks. A simulation study was performed with the proposed control system in place to show the effect of the proposed control system.

5.4.2 Controllability

Definition 7: Controllability means that a system state can be pulled into a state in a predefined desired region from any other state by unconstrained control signals in a finite interval of time.

Definition 8: A system switching event set is defined as a set of transitions between two consecutive abstract network states:

$E = E_{uc} \cup E_c = PHASE \rightarrow PHASE$

where $E_{uc} \cap E_c = \varnothing$ and E_c = {Bi-State → Bi-State, Bi-State → Congestion, Congestion → Bi-State, Congestion → Congestion}

Lemma 2: E_c is controllable

Proof of Lemma 2: When the controller sees a Bi-State or Congestion, a high congestion fee will be charged according the controller design. Because of the dynamics equations of agents, most users, including all heavy users will reduce the sending rate to a low level. That is, an event $e \in E_c$ is excludable. Thus, E_c is controllable.

Lemma 3: E_{uc} is uncontrollable

Proof of Lemma 3: Free Way and Saturated are always allowed. Thus, E_{uc} is uncontrollable.

Theorem 1: The cable network system is controllable with respect to the system switching event set, *E*.

Proof of Theorem 1: According to the controllability theorem stated in [85] and Lemma 2 and 3, the cable network system is controllable with respect to system switching event set, *E*. Intuitively, the cable network system only experiences bi-state or congestion ephemerally and the interval of this brief period can be adjusted by the cable network operator.

Theorem 2: The cable network system is controllable with respect to the subsystem 1.

Proof of Theorem 2: Demand theory states that price can restrict the use of network resources. Therefore, high dynamic prices can restrict the use of network resources and low dynamic prices can encourage the use of network resources. Specifically, from the dynamic equation of agents (equation (17)), agents increase sending rate, $x_{i,k}$, when the price, p_k, is lower than the marginal utility and decrease sending rate, $x_{i,k}$, when the price, p_k, is higher than the marginal utility. Thus, cable network system is controllable with respect to subsystem 1.

5.4.3 Stability

Definition 9: Stability in the sense of congestion is defined as no long-period congestion or bi-state.

Definition 10: Stability in the sense of system efficiency is defined as keeping subsystem 1 along a sub-efficiency frontier, i.e. a sliding surface.

Lemma 4: $f_1(y, p) \geq 0$ if $s - s_{th} \leq 0$ and $f_1(y, p) < 0$ if $s - s_{th} > 0$

Proof of Lemma 4: The sign of $f_1(y, p)$ is determined by agent dynamics.

Case 1: If $s - s_{th} \leq 0$, there is no control signal, i.e. no dynamic price is charged, and agents tend to increase the sending rate, which leads to $f_1(y, p) \geq 0$.

Case 2: If $s - s_{th} > 0$, there is a control signal, i.e. dynamic price is activated, and agents tend to decrease the sending rate, which leads to $f_1(y, p) < 0$.

Theorem 3: The cable network is stable in the sense of congestion

Proof of Theorem 3: Theorem 1 also states that long-period congestion or bi-state will not occur, i.e. the cable network is stable in the sense of congestion. A short period of congestion can be thought of as a disturbance.

Theorem 4: The cable network is stable in the sense of system efficiency.

Proof of Theorem 4: Equivalently, Theorem 4 states that subsystem 1 can be kept in a neighborhood of the sliding surface, i.e. a sub-efficiency frontier. Here, the equilibrium is a surface instead of a point as in a traditional controlled system.

Let a Lyapunov function be

$$V = (s - s_{th})^2/2 \tag{24}$$

where, V is continuously positive definite. Then, using (14), we obtain

$$dV/dt = (s - s_{th})\, ds/dt = (s - s_{th}) f_1(y, p) \tag{25}$$

According to Lemma 4, $dV/dt \leq 0$. Based on a stability theorem stated in [43], subsystem 1 is uniformly stable. Here, V can be interpreted as an energy function and $dV/dt \leq 0$ means that this energy always tends to decrease with time. Figure 19 illustrates this stability for a cable network with two agents.

5.4.4 Sub-efficiency of Cable Networks

Theorem 5: The proposed cable network control is sub-efficient.

Proof of Theorem 5: As shown in Figure 19 with two users, the Pareto frontier separates the $\sigma = 1$ and $\sigma = 2$ regions. The hyper-plane (or line) $\sum_{i=1}^{n} y_i = s_{th}$ is a sub-efficiency frontier the control system wants to achieve. There is an efficiency loss equal to the area between the Pareto frontier and the sub-efficiency frontier. Thus, the cable network control system makes the cable network sub-efficient.

Given the sub-efficiency frontier, the controller encourages light users to approach the normal use indicated by the point on the Pareto frontier where $y_1 = y_2 = s_{max} / 2$ in Figure 19. This controller design also aims to protect the light users through limiting the excessive use of bandwidth even though there is some bandwidth available for heavy users.

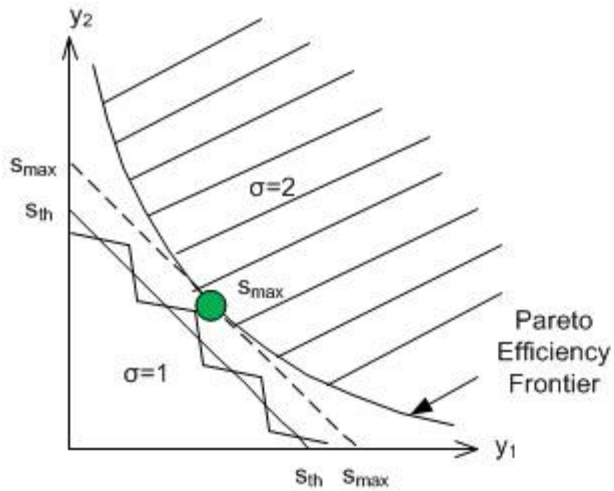

Figure 19: Stability of cable network with two agents

5.4.5 Voluntary Participation

Theorem 6: If a cable network with flat-pricing satisfies the property of voluntary participation, a cable network with the proposed control also does.

Proof of Theorem 6:

Case 1: For light and normal users, there is no change. From equation (12), $u_i(y_i/cap) > 0$ if $y_i > 0$ for all light and normal users.

Case 2: For heavy users, there is no change when they choose light use or normal use (Normal use is the prescribed or provisioned bandwidth use). From equation (12), $u_i(y_i/cap) > 0$ if $y_i > 0$ for all heavy users with light use or normal use.

Subscription is based on normal use. Thus, if a cable network with flat-pricing satisfies the property of voluntary participation, a cable network with the proposed control also does.

5.4.6 Incentive Compatibility

Theorem 7: The proposed cable network control is incentive compatible.

Proof of Theorem 7: Being incentive compatible is also called being strategy proof. When $s_{th} < s$ without congestion or the system is in congestion, an additional significant fee (see controller design) is charged for bandwidth use. Assume that a heavy user i fakes his/her valuation function, $v_i(y_i)$, such that the marginal valuation, $dv_i(y_i)/dy_i$, is greater than the current price and that the actual marginal valuation is less than or equal to the current price. Then, this user suffers a utility loss (i.e. negative utility, $u_i(y_i/cap) < 0$) while transmitting. Since a rational user does not act to get a negative utility, this user will not falsify his/her valuation function. Therefore, the proposed cable network control is incentive compatible. This implies that local utility maximizations of agents lead to the cable network utilization sub-efficiency set by the mechanism designer.

5.4.7 Simulation Study of the Cable Network Control System

An NS2 simulation is conducted to demonstrate the effectiveness of the control system of cable networks. In this simulation, it is assumed that CMs are homogeneous heavy users. Figure 20 shows the cable network throughput without the proposed control system while CMs increase rates with time. The cable network gets congested starting from around 13 seconds; and the faster the CMs sending rate, the more severe the congestion is. Figure 21 shows the cable network throughput with CMs manipulating rates with time when the proposed control system is in place. With the proposed control system, the cable network gets controlled and keeps a throughput of about 0.95 Mbps, which is close to the peak throughput of about 1 Mbps shown in Figure 20. The cable network congestion is effectively controlled by the proposed cable network control system and the cable network is proportionally fair at the same time.

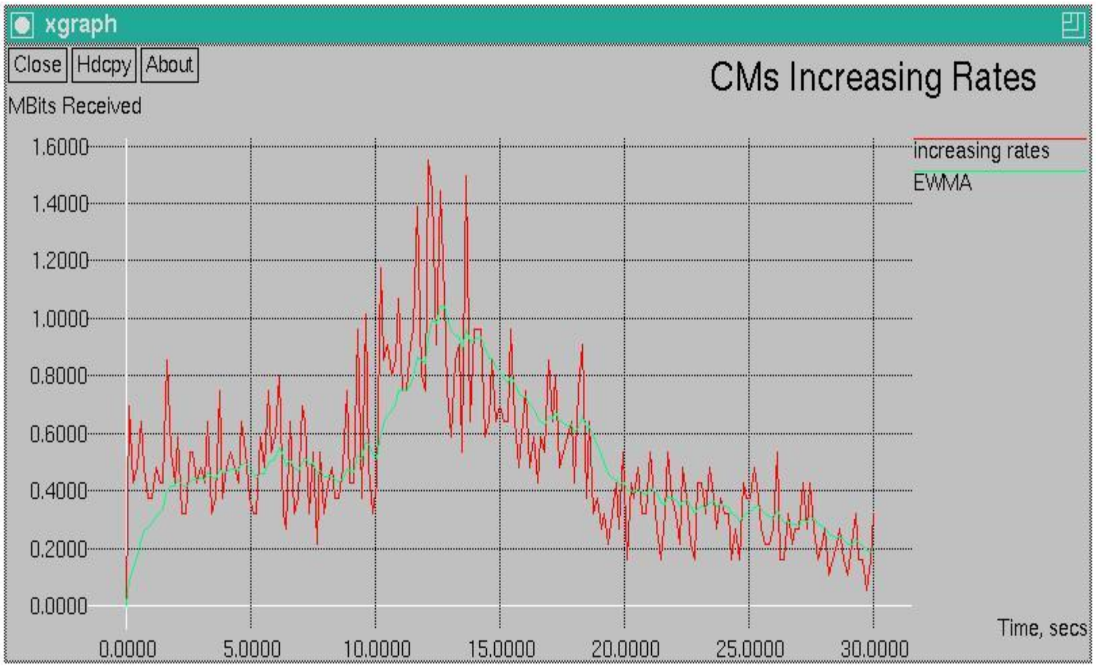

Figure 20: Cable network throughput without the proposed control system

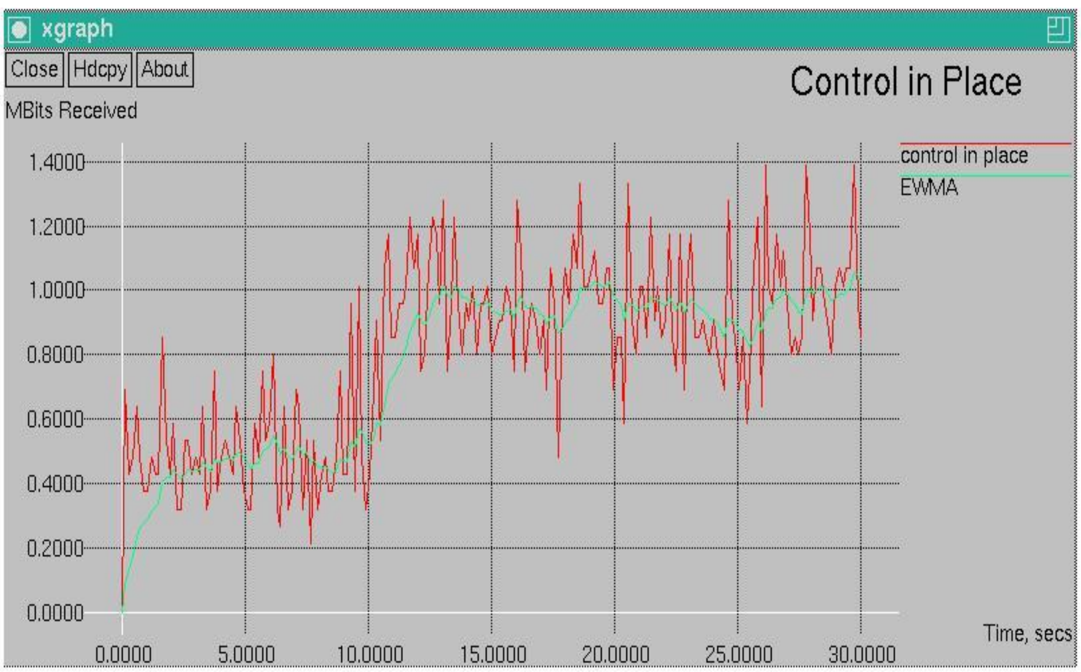

Figure 21: Cable network throughput with the proposed control system

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

This research proposes a feedback control system to solve bandwidth allocation problems, such as congestion and fairness, for cable networks. Modeling and analysis of cable networks, observer design of cable networks, and controller synthesis of cable networks have been conducted. The prescribed QoS may be unrealizable with existing controls, but the proposed feedback control system for cable networks has desirable properties such as controllability, stability, sub-efficiency, voluntary participation, incentive compatibility, simplicity, and practicability. A simulation study shows that the proposed control system effectively keeps the cable network state in a desired region. This research is expected to have three impacts on cable networks: improved bandwidth utilization, fairness, and security strength.

Game theory, combined with system theory, has been used to model cable networks. In addition to traffic flow classification and utilization smoothing, a hidden Markov model has been applied to identify the network states. Controller design has been done by designing new dynamic games with the methodology of mechanism design. The control objectives are achieved by distributive agent-based control, and control techniques, such as switched-system control and sliding mode control, have been used to cope with the model uncertainties of cable networks.

One future work could be to extend the current work to inter-networks, where ISPs act as autonomous agents, to make QoS truly attainable while the best-effort nature of packet-switched computer networks is still retained. Max-min fairness requires complicated allocation rules, payment rules, and communication schemes and will be addressed in future work. It would be very valuable to test the proposed solution on a real-world cable network in addition to theoretical analyses and simulation studies.

REFERENCES

REFERENCES

[1] Cable Television Laboratories, Inc., "Data-Over-Cable Service Interface Specifications DOCSIS 3.0: MAC and Upper Layer Protocols Interface Specification," http://bradyvolpe.com/library/CM-SP-MULPIv3.0-I13-100611.pdf, 2010. [cited April 3, 2011].

[2] E. N. Barron, Game Theory: an Introduction. Hoboken, NJ: John Wiley & Sons Inc., 2008.

[3] J. Watson, Strategy, An Introduction to Game Theory, 2nd ed.. New York, NY: W. W. Norton & Company, 2007.

[4] Y. Shoham, and K. Leyton-Brown, Multiagent Systems: Algorithmic, Game-Theoretic and Logical Foundations. New York, NY: Cambridge University Press, 2009.

[5] J. Chen and et al, “Game theoretical approach for channel allocation in wireless sensor and actuator networks,” IEEE Transactions on Automatic Control, vol.56, no. 10, pp 2332-2344, 2011.

[6] S. Sengupta, M. Chatterjee, and K. A. Kwiat, “A game theoretic framework for power control in wireless sensor networks,” IEEE Transactions on Computers, vol. 59, no. 2, pp. 231-242, 2010.

[7] Y. Chevaleyre and et al., “Issues in multiagent resource allocation,” Informatica. vol. 30, pp. 3-31, 2006.

[8] J. Martin, and J. Westall, “A simulation model of the DOCSIS protocol,” Simulation. vol. 83, pp. 139-155, 2007.

[9] W. Lee, K. Chung, and K. Chu, and J. Pan, “DOCSIS performance analysis under high traffic conditions in the HFC networks,” IEEE Transactions on Broadcasting. vol. 52, pp. 21-30, 2006.

[10] W. K. Kuo, S. Kumar, C.-C. Jay Kuo, “Improved priority access, bandwidth allocation and traffic scheduling for DOCSIS cable networks,” IEEE Transactions on Broadcasting. vol. 49, pp. 371-382, 2003.

[11] J. Min, Cross-user subsidy in residential broadband service. MS Thesis. http://dspace.mit.edu/bitstream/handle/1721.1/34538/70963145.pdf?sequence=1, 2005. [cited May 18, 2011].

[12] R. Edell and P. Varaiya, “Providing internet access: what we learn from INDEX,” IEEE Network. vol. 13, pp. 18-25, 1999.

REFERENCES (continued)

[13] Mao-Bin Hu, Wen-Xu Wang, Rui Jiang, Qing-Song Wu, and Yong-Hong Wu, “Phase transition and hysteresis in scale-free network traffic,” Physical Review E, vol. 75, pp. 036102, 2007.

[14] Y. Tretyakov, H. Takayasu, and M. Takayasu, “Phase transition in computer network model,” Physica A, vol. 253, pp. 315-22, 1998.

[15] T. Ohira and R. Sawatari, “Phase transition in a computer network traffic model,” Physical Review E, vol. 58, pp. 193-5, 1998.

[16] R. V. Solé and S. Valverde, “Information transfer and phase transitions in a model of Internet traffic,” Physica A, vol. 289, pp. 595-605, 2001.

[17] G. Mukherjee and S. S. Manna, “Phase transition in a directed traffic flow network,” Physical Review E, vol. 71, pp. 066108, 2005.

[18] L. Kocarev and G. Vattay, Eds., Complex Dynamics in Communication Networks, New York, NY: Springer-Verlag, 2005.

[19] R. R. Kompella and Cristian Estan, “The power of slicing in Internet flow measurement,” *IMC '05 Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement*, pp. 105-118, 2005.

[20] C. Scoglio, C. Bruni, G. Koch, and S. Sutrave, “Estimation of data traffic from aggregate measurements”, Mathematical and Computer Modeling, vol. 52, pp. 837-853, 2010.

[21] Thuy T.T. Nguyen and G. Armitage, "A survey of techniques for Internet traffic classification using machine learning," IEEE Communications Survey & Tutorial, vol. 10, no. 4, pp. 56-76, 2008.

[22] J. E. B. Maia and R. H. Filho, "Internet Traffic Classification using a Hidden Markov model," 2010 10th International Conference on Hybrid Intelligent Systems, pp. 37-42, 2010.

[23] K. Xu, Z.-L. Zhang, and S. Bhattacharyya, "Internet traffic behavior profiling for network security monitoring," IEEE/ACM Transactions on Networking, vol. 16, no. 6, pp. 1241-52, 2008.

[24] X. Tian, Q. Sun, X. Huang,and Y. Ma, "A Dynamic Online traffic classification methodology based on data stream mining," 2009 World Congress on Computer Science and Information Engineering, pp. 298-302, 2009.

[25] V. P. de A Ribeiro, R. H. Filho, J. E. B. Maia, "Online traffic classification based on sub-flows," 2011 IFIP/IEEE International Symposium on Integrated Network Management, pp. 415-21, 2011.

[26] Dainotti, W. Donato, A. Pescape, and P. S. Rossi, "Classification of network traffic via packet-level hidden markov models," IEEE GLOBECOM, 2008.

[27] C. Wright, F. Monrose, and G. Masson, "Hmm profiles for network traffic classification," in VizSEC/DMSEC, pp. 9–15, 2004.

[28] R. O. Duda, P. E. Hart, and D. G. Stock, Pattern Classification, 2nd ed., New York, NY: John Wiley & Sons, 2000.

[29] L. Rabiner and B.-H. Juang "Theory and Implementation of Hidden Markov Models", Fundamentals of Speech Recognition, pp. 321-388, New York, NY: Prentice-Hall, 1993.

[30] N. Ghanem, D. DeMenthon, D. Doermann, and L. Davis, "Representation and recognition of events in surveillance video using petri nets", Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, pp. 112-120, 2004.

[31] H. Gao, A. Jasti, R. Pendse, and M. E. Sawam, "Survey of recent advances in control of computer networks: models, approaches, and autonomic applications", *Unpublished*.

[32] Y. H. Aoul, A. Mehaoua, and C. Skianis, "A fuzzy logic-based AQM for real-time traffic over Internet," Computer Networks, vol 51 no. 16, pp. 4617-33, 2007.

[33] O,.P. Kreidl and T.M. Frazier, "Feedback control applied to survivability: a host-based autonomic defense system," IEEE Transactions on Reliability, vol. 53, no. 1, pp. 148-66, 2004.

[34] G. Kaiser, J. Parekh, P. Gross, and G, Valetto, "Kinesthetics eXtreme: an external infrastructure for monitoring distributed legacy systems," Proceedings of the Autonomic Computing Workshop: 5fth Annual International Workshop on Active Middleware Services, pp. 22-30, 2003.

[35] X. Shi et al, "Pattern recognition based adaptive real-time scheduling," International Conference on Machine Learning and Cybernetics, pp. 3160-3166, 2003.

[36] H.-P. Huang, J.-Y. Lee, and W.-H. Cheng, "Network flow control based on fuzzy colored timed Petri net model in the OSI session layer," International Journal of Fuzzy Systems, vol. 4, no. 4, pp. 883-95, 2002.

REFERENCES (continued)

[37] H. Yu and C. G. Cassandras, "Perturbation analysis of communication networks with Feedback Control Using Stochastic Hybrid Models," Nonlinear Analysis, vol. 65, no. 6, pp. 1251-80, 2006.

[38] D. Liberzon, Switching in Systems and Control, Boston, MA: Birkhauser, 2003.

[39] E. Lehtihet, H. Derbel, N. Agoulmine, Y. Ghamri-Doudane, and S. Meer, "Initial approach toward self-configuration and self-optimization in IP network," 8th International Conference on Management of Multimedia Networks and Services, pp. 371-82, 2005.

[40] T. F. Abdelzaher, K. G. Shin, and N. Bhatti, "Performance guarantees for Web server end-systems: a control-theoretic approach," IEEE Transactions on Parallel and Distributed Systems, vol. 13, no. 1, pp. 80-96, 2002.

[41] J. F. Kurose and K. W. Ross, Computer Networking: A Top-Down Approach Featuring the Internet, 3rd ed., New Jersey: Pearson Education Inc, 2005.

[42] G. Box and A. Luceño, Statistical Control by Monitoring and Feedback Adjustment, John Wiley & Sons, inc., 1997.

[43] H. Khalil, Nonlinear Systems, Prentice Hall, 3rd edition, Upper Saddle River, NJ: Prentice Hall, 2002.

[44] S. R. Williams, Communication in mechanism design: a differential approach, Cambridge University Press, 2008.

[45] D. Garg and Y. Narahari, "Mechanism design for single leader Stackelberg problems and application to procurement auction design," IEEE Transactions on Automation Science and Engineering, vol. 5, no. 3, pp. 377-393, 2008.

[46] R. K. Dash, N. R. Jennings, and D. C. Parkes "Computational-Mechanism Design: A Call to Arms," IEEE Intelligent Systems, pp. 40-47, vol. 18, no. 6, 2003.

[47] J. W. Huang and V Krishnamurthy, "Application of mechanism design in opportunistic scheduling under cognitive radio systems," 2009 Second International Conference on Advances in Mesh Networks, pp. 50-54, 2009.

[48] L. Tang, Q. Chen, X. Zhang, X. Zeng, and Y. Li, "Call admission control for multiservice radio access network using mechanism design," WiCOM '08. 4th International Conference on Wireless Communications, Networking and Mobile Computing, pp. 1-5, 2008.

REFERENCES (continued)

[49] Man Hon Cheung, Amir-Hamed Mohsenian-Rad, Vincent W.S. Wong, and Robert Schober, "Random access protocols for WLANs based on mechanism design," Proceedings of IEEE International Conference on Communications (ICC), 2009.

[50] F. Fu and M. van der Schaar, "Noncollaborative resource management for wireless multimedia applications using mechanism design," IEEE Transactions on Multimedia, vol. 9, no. 4, pp. 851-868, 2007.

[51] J. Zhang, Z. Zhao, Y. Gong and M. Ma, "A decentralized incentive mechanism for P2P file sharing systems," International Conference on Information, Communications & Signal Processing, pp. 1-5, 2007.

[52] C. Ko and H. Wei, "On-demand resource-sharing mechanism design in two-tier OFDMA femtocell networks," IEEE Transactions on Vehicular Technology, vol. 60, no. 3, pp. 1059-1071, 2011.

[53] A. K. Chorppath and T. Alpcan, "Mechanism design for energy efficiency in wireless networks," The 7th International Workshop on Resource Allocation and Cooperation in Wireless Networks (RAWNET/WNC3), pp. 389-394, 2011.

[54] S. Kunniyur, and R. Srikant, "End-to-end congestion control schemes: utility functions, random losses and ECN marks," IEEE/ACM Transactions on Networking, vol. 11, no.5, pp. 689-702, 2003.

[55] W. Wang, X. Li, and Z. Sun, "Design differentiated service multicast with selfish agents," IEEE Journal on Selected Areas in Communications, vol. 24, no. 5, pp. 1061-1073, 2006.

[56] Duy H. N. Nguyen and Tho Le-Ngoc, "Multiuser downlink beamforming in multicell wireless systems: a game theoretical approach," IEEE Transactions on Signal Processing vol. 59, no. 7, pp. 3326-3338, 2011.

[57] P. Gevros, J. Crowcroft, P. Kirstein, and S. Bhatti, "Congestion control mechanisms and the best effort service model," IEEE Network, pp. 19-26, 2001.

[58] S. Floyd and K. Fall, "Promoting the use of end-to-end congestion control in the internet , IEEE/ACM Transactions on Networking," vol. 7, no. 4, pp. 458-472, 1999.

[59] R. Cocchi, S. Shenker, D. Estrin, and L. Zhang, "Pricing in computer networks: motivation, formulation, and example," IEEE/ACM Transactions on Networking, vol. 1, no. 6, pp. 614-627, 1993.

REFERENCES (continued)

[60] N. AbuAli, M. Hayajneh, and H. Hassanein, "Congestion-based pricing resource management in broadband wireless networks," IEEE Transactions on Wireless Communications, vol. 9, no. 8, pp. 2600-2610, 2010.

[61] E. W. Fulp and D. S. Reeves, "Distributed network flow control based on dynamic competitive markets," Proceedings International Conference on Network Protocol (ICNP'98), pp. 119-128, 1998.

[62] J. Hou, J. Yang, and S. Papavassiliou, "Integration of pricing with call admission control for wireless networks," in Proc. IEEE VTC 01, vol. 3, pp. 1344-1348, 2001.

[63] T. H. Batubara, C. Y. Huat, and M. Singh, "On modeling the effect of peak-load pricing mechanism to the telecommunication traffic," Proceedings of VTC , pp.1-5, 2010.

[64] T. Moscibroda and S. Schmid, "On mechanism design without payments for throughput maximization," IEEE INFOCOM 2009, pp. 972-980, 2009.

[65] L. Chen, S. H. Low, and J. C. Doyle, "Random access game and medium access control design," IEEE/ACM Transactions on Networking, vol. 18, no. 4, pp. 1303-1316, 2010.

[66] C. Singh, A. Kumar and R. Sundaresan, "Uplink power control and base station association in multichannel cellular networks," International Conference on Game Theory for Networks, pp. 43-51, 2009.

[67] C. Long, Q. Zhang, B. Li, H. Yang, and X. Guan, "Non-Cooperative power control for wireless ad hoc networks with repeated games," IEEE Journal on Selected Areas in Communications, vol. 25, no. 6, pp. 1101-1112, 2007.

[68] D. K. Sanyal, M. Chattopadhyay, and S. Chattopadhyay, "Improved performance with novel utility functions in a game-theoretic model of medium access control in wireless networks," IEEE Region 10 Conference, pp. 1-6, 2008.

[69] J. Price, P. Nuggehalli, T. Javidi, "Incentive compatible MAC-layer QoS design, Proceedings of IEEE Consumer Communications and Networking Conference," pp. 586-590, 2008.

[70] J. Lee, W. Liao, and M. C. Chen, "An incentive-based fairness mechanism for multi-hop wireless backhaul networks with selfish nodes," IEEE Transactions on Wireless Communications, vol. 7, no. 2, pp. 697-704, 2008.

[71] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: Shadow prices, proportional fairness and stability," Journal of Operation Research Society, vol. 49, no. 3, pp. 237-252, 1998.

REFERENCES (continued)

[72] Y. Yang, Z. Cao, M. Tan, and J. Yi, "Fairness and dynamic flow control in both unicast and multicast architecture networks," IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews, vol. 37, no. 2, pp. 206-212 , 2007.

[73] D. P. Bertsekas, Nonlinear Programming, Belmount, MA: Athena Scientific, 1995.

[74] T. Cui, L. Chen, S. H. Low, "A game-theoretic framework for medium access control," IEEE Journal on Selected Areas in Communications, vol. 26, no. 7, pp. 1116-1127, 2008.

[75] M. Rasti, A. R. Sharafat, and B. Seyfe, "Pareto-efficient and goal-driven power control in wireless networks: a game-theoretic approach with a novel pricing scheme," IEEE/ACM Transactions on Networking, vol. 17, no. 2, pp. 556-569, 2009.

[76] H. Inaltekin and S. B. Wicker, "The analysis of Nash equilibria of the one-shot random-access game for wireless networks and the behavior of selfish nodes," IEEE/ACM Transactions on Networking, vol. 16, no. 5, pp. 1094-1107, 2008.

[77] K. Akkarajitsakul, E. Hossain, D. Niyato, and D. I. Kim, "Game theoretic approaches for multiple access in wireless networks: a survey," IEEE Communications Surveys & Tutorials, vol. 13, no. 3, pp. 372-395, 2011.

[78] G. Kotsalis and J. S. Shamma, "Robust synthesis in mechanism design," 49th IEEE Conference on Decision and Control, pp. 225-230, 2010.

[79] P. B. Luh, S. C. Chang and T. S. Chang, "Solutions and properties of multistage Stackelberg games, Automatica, " vol. 20, no. 2, pp. 251-256, 1984.

[80] T. Alpcan, L. Pavel and N. Stefanovic, "A control theoretic approach to noncooperative game design," Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, pp. 8575-8580, 2009.

[81] Mehran S. Fallah, "A puzzle-based defense strategy against flooding attacks using game theory," IEEE Transactions on Dependable and Secure Computing, vol. 7, no. 1, pp. 5-19, 2010.

[82] N. Gregory Mankiw, Principles of Microeconomics, 6th ed., South-Western Cengage Learning, Mason, OH, USA, 2012.

[83] T. M. Humphrey, "Marshallian cross diagrams and their uses before Alfred Marshall: the origins of supply and demand geometry," Economic Review, vol. 78, no. 2, pp. 3-23, 1992.

[84] Debian, "Source package: ns2 (2.35~rc10+dfsg-1)," http://packages.debian.org/source/sid/ns2, [cited January 16, 2012].

[85] C. G. Cassandras and S. Lafortune, Introduction to Discrete Event System. New York: Springer, 1999.

APPENDIX

APPENDIX

KEY TERMINOLOGIES

Adaptation: Regulation and reorganization according to the environmental or internal system changes/patterns

Agent: An individual or entity who makes independent strategic decisions. In my research, agents include the cable network subscribers and a social planner (ISP will delegate this social planner)

Automation: Self-management or self-operation

Autonomic: An autonomic system possesses the following three properties:

- Environment awareness
- Adaptation
- Automation

Computer network models: A network model defines the states (micro or macro) and catches the dynamics of computer networks

Control of network: A methodology of using control theory to design and implement autonomic networking

Distributed & agent-based control: The goal of the system controlled is realized by distributed agents by message communication on networks

Environment Awareness: Ability to sense the environmental and internal system changes

Features of networks: Observable network states

Macro system states: Conditions of the whole system

Micro system states: Conditions of agents

APPENDIX (continued)

Network observer: A mechanism to estimate network hidden (unobservable) states

Price incentive: A monetary reward or penalty to spur or to stop some actions

Social welfare: Social welfare functions are metrics to measure the performance of the whole system. Two typical metrics used are utilitarian (a special case is utilization) and egalitarian functions

Utility: A utility of an agent specifies the preference order of his/her possible actions through valuing micro system states