2014

Dynamic Pricing in Heterogeneous Wireless Cellular Networks

David Shrader
Nova Southeastern University, dshrader@nova.edu

Dynamic Pricing in Heterogeneous Wireless Cellular Networks

by

David Shrader

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Computer Information Systems

Graduate School of Computer and Information Sciences

Nova Southeastern University

2014

We hereby certify that this dissertation, submitted by David Shrader, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

Gregory E. Simco, Ph.D. Chairperson of Dissertation Committee	Date
Francisco J. Mitropoulos, Ph.D. Dissertation Committee Member	Date
Sumitra Mukherjee, Ph.D Dissertation Committee Member	Date

Approved:

Eric S. Ackerman, Ph.D. Dean, Graduate School of Computer and Information Sciences	Date

Graduate School of Computer and Information Sciences
Nova Southeastern University

2014

An Abstract of a Dissertation Submitted to Nova Southeastern University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Dynamic Pricing in Heterogeneous Wireless Cellular Networks

by
David Shrader
August 2014

Smart communications devices are giving users instant access to applications that consume large amounts of data. These applications have different requirements on the network for delivery of data. In order to support these different applications, operators are required to support multiple service classes.

Given the regulatory and technology constraints and the relatively high cost associated with wireless spectrum licensing and utilization, demand will exceed supply leading to congestion and overload conditions. In addition to new broadband radio technologies offering higher data rates, operators are looking at deploying alternate heterogeneous technologies, such as WLAN, to provide additional bandwidth for serving customers. It is expected that this will still fall short of providing enough network resources to meet the ITU requirement for 1% new call blocking probability. An economic mechanism that offers incentives to individuals for rational behavior is required in order in order to reduce the demand for network resources and resolve the congestion problem.

The research in this dissertation demonstrates that the integration of a dynamic pricing with connection admission control mechanism for an operator deploying cooperative heterogeneous networks (e.g., LTE and WLAN) offering multiple QoS service classes reduces the new call blocking probability to the required 1% level.

The experimental design consisted, first, of an analytical model of the CAC algorithm with dynamic pricing in a heterogeneous environment. The analytical model was subsequently validated through discrete-event simulation using Matlab.

Acknowledgements

Completion of this dissertation process has been a difficult task to accomplish. I appreciate the support of both the faculty and administration of the Graduate School of Computer and Information Sciences for their support throughout. I would like to specially acknowledge the efforts of Dr. Gregory Simco for sticking with me through the last three years as I persevered to return to identify and refine my research problem, define an appropriate methodology and proposal and finally to guide me through the simulation problem solving and analysis prior to the final completion of the report.

My good friend, Dr. Raquel Morera, has been quite supportive for motivation and a sounding board for ideas, spending large amount of time on the phone or over dinner discussing what I meant when I wrote something unclear down on paper.

Finally, I would like to acknowledge the support of coworkers and friends at Telcordia and Ericsson, especially my direct supervisor, David Gorton, for putting up with me during this work.

Table of Contents

List of Tables

Tables

List of Figures

Figures

Chapter 1

Introduction

Introduction

Smart communications devices are giving users instant access to applications that consume large amounts of data. Cisco (2013) projects that mobile data traffic will exceed fixed network data traffic and experience a compound annual growth rate of 66% between 2012 and 2017 due largely to internet video and file sharing. Other application categories include web, email and data; video and voice over IP calling, and online gaming. Given the regulatory and technology constraints and the relatively high cost associated with wireless spectrum licensing and utilization, demand will exceed supply leading to congestion and overload conditions (AbuAli, Hayajneh, & Hassanein, 2010; Al-Manthari, Nasser, Ali, & Hassanein, 2009; Al-Manthari, Nasser, & Hassanein, 2011; Courcoubetis & Weber, 2003; Ekstrom, 2009).

The work described in this paper successfully introduces a dynamic pricing algorithm into a channel allocation and call admission control function in a cooperative heterogeneous WiFi and LTE network to reduce the call blocking probability, one measure of quality of service, to below 1%. The work extends the original model proposed by Hou, Yang, and Papavassiliou (2002) to a multi-service, heterogeneous network involving two radio access technologies.

Hou, Yang, and Papavassiliou (2002) reasoned that when a user looks to maximize his/her own satisfaction when consuming a limited resource such as network bandwidth, the resulting aggregate load on the network will be greater than available resources and the total usefulness to the community is decreased. Since the total demand is the result of many individual decisions, a mechanism that offers incentives to individuals for rational behavior can affect the demand for network resources in order to resolve the congestion problem. To encourage rational user behavior, Hou et al. introduced a dynamic pricing function that prices the network service in a manner that requires the users to self-regulate based on their willingness to pay for the service, resulting in fewer requests being blocked by the network with probability less than 1% even during peak usage times.

In order to provide quality of service guarantees to the subscriber, one of the controls employed by the network is a channel allocation and call admission control function (Ahmed, 2005; Courcoubetis & Weber, 2003; Ekstrom, 2009; Soldatos, Vayias, & Kormentzas, 2005). This function, known generically as CAC, only accepts service requests when there is adequate network bandwidth available and reserves the resources necessary for the service up front prior to user data transmission. If there is insufficient bandwidth available, the service request is rejected. This ensures that congestion in the network does not affect the quality of service offered to the subscriber when it is accepted into the network.

The key innovation provided by Hou, Yang, and Papavassiliou (2002) is the introduction of a dynamic pricing function to the CAC algorithm. Hou et al. prove that there is an optimal arrival rate of traffic for any network and that congestion and the associated call blocking occurs when the demand is greater than ability of the network to handle or, in other words, the actual arrival (i.e., demand) rate is greater than the optimal arrival rate for the network. Then, assuming users are price sensitive, the logic is that by raising the price dynamically when the arrivals exceed the optimum, enough users will choose not to enter the system and the arrival rate is kept below the optimum leading to a congestion free system. The coupling with CAC, allows the already accepted users to continue their service unhindered by increasing numbers of arrivals. The remainder of this section provides an overview of historical strategies for handling network growth and the application of dynamic pricing with CAC to today's wireless network environment.

From an economist's view, the usefulness of network services grows with its size which, in turn, has driven this large increase in demand (Courcoubetis & Weber, 2003). Courcoubetis and Weber outline two key alternatives to meet this demand.

The first is to over-engineer the network to provide more capacity than required by users, thus guaranteeing the ability for the network to satisfy the user demand. This was the general approach used for the Internet growth phase of the 1990's. In a fixed network environment, this is achieved by deploying more physical facilities (wires and cables).

In the wireless domain, however, the wireless frequency spectrum is considered a social resource and is licensed to operators in limited segments within which they can offer services. Wireless technologies have been developed with increased bandwidth to allow the operator to offer data services with higher bandwidth within their licensed spectrum. WiMAX (IEEE, 2009) and LTE (3GPP, 2010a) are mobile broadband technologies that require licensed spectrum and the network operators are limited in a given area by the amount of bandwidth their licensed spectrum will support. In the unlicensed spectrum domain, Wireless LAN technology (IEEE, 2007), also known as WLAN, is planned for use to extend the operator service coverage (3GPP, 2009, 2010b, 2011). Some operators such as KT Corporation in Korea are deploying a three-tier network involving the current Wideband CDMA with both WiMAX and WLAN to provide high speed coverage for users (Choi, Ji, Park, Kim, & Silvester, 2011).

Given the regulatory and technology constraints and the relatively high cost associated with spectrum licensing and utilization, over-engineering the whole network is not a viable solution and demand will exceed supply leading to congestion and overload conditions (AbuAli et al., 2010; Al-Manthari et al., 2009; Al-Manthari et al., 2011; Courcoubetis & Weber, 2003; Ekstrom, 2009).

Courcoubetis and Weber (2003) offer a second strategy for operators to meet the increased demand. Equip the network with controls that are in effect at all times allowing different customers access to the specific services that have value to them and they are willing to pay for. Courcoubetis and Weber note in economic terms, the social value of the network is increased when users can select specific services appropriate for their needs and that for this service differentiation to be effective, attractive prices are required for customers to utilize those services. In this way, pricing becomes a control mechanism that can be used by an operator to generate the appropriate demand.

The key differentiation for services is provided by the quality of service associated with the transmission channel, known as a bearer (Ekstrom, 2009). The application categories introduced above for internet video, file sharing, email and data, video and voice over IP calling, and online gaming all require different performance characteristics for a satisfactory user experience in terms of bit rates, packet delays, and packet error loss rate. For heterogeneous network and multi-vendor operation, 3GPP has defined a set of nine QoS classes to characterize the network-level treatment (e.g., scheduling, queuing, link-layer configuration) provided to traffic associated with a particular service.

In order to provide service differentiation, the operator specifies the QoS class, maximum bit rate and guaranteed bit rate for each service that it offers to subscribers. In order to provide for subscriber differentiation, a priority can also be assigned to the service that is used to determine how to handle requests for service when the network is in a congested state.

In order to provide quality of service guarantees to the subscriber, one of the controls employed by the network is a channel allocation and call admission control function (Ahmed, 2005; Courcoubetis & Weber, 2003; Ekstrom, 2009; Soldatos et al., 2005). This function, known generically as CAC, only accepts service requests when there is adequate network bandwidth available and reserves the resources necessary for the service up front prior to user data transmission. If there is insufficient bandwidth available, the service request is rejected. This ensures that congestion in the network does not affect the quality of service offered to the subscriber when it is accepted into the network.

The mobility aspect of subscribers in a wireless network introduces an added complexity. When a subscriber moves into a new cell, there is a potential that the new cell will be congested and the subscriber's session will be dropped due to lack of available bandwidth.

Network operators engineer their network to meet the standard set by the International Telecommunications Union for normal operation such that the probability that a subscriber session will be dropped is less than 1% (ITU-R, 2000, 2006).

In a comprehensive survey of CAC methods for wireless networks, Ahmed (2005) identifies two key quality of service parameters related to session blocking prevalent in the CAC research: new call blocking probability and handoff call blocking probability. Ahmed notes that a common characteristic of all CAC methods is that a reduction in the handoff call blocking probability is generally accompanied by an increase in the new call blocking probability. Hou, Yang, and Papavassiliou (2002) studied a variety of CAC schemes and observed that with an increase in the number of service requests by consumers there is in an increase in the overall call blocking probability. During peak usage times, the call blocking probability approached 8% far exceeding the standard set by the International Telecommunications Union.

While CAC allows the network to provide a guaranteed quality of service to subscribers, it does so by blocking requests for service when the network is congested (i.e., all bandwidth is allocated to current users). In economic terms, when a resource shared by all users is degraded by the introduction of one user as in a congested network, this is known as a negative market externality (Courcoubetis & Weber, 2003; Hou et al., 2002). For maximum benefit of the shared resource, the externality must be internalized. That is, for rational user behavior, the congestion externality must be assigned a price. This additional price provides users incentive to use the network services rationally.

To encourage rational user behavior, Hou et al. introduced a dynamic pricing function combined with CAC. The dynamic pricing function prices the network service in a manner that requires the users to self-regulate based on their willingness to pay for the service, resulting in fewer requests being blocked by the network with probability less than 1% even during peak usage times.

This paper focuses on the prevention of congestion, in the form of call blocking probability less than 1%, for a multi-class service network using heterogeneous wireless technologies such as the operators are planning with a combination of LTE and WLAN coverage.

Problem Statement

When demand exceeds available resources, the wireless network is in a congestion condition and user data needs in the form of Quality of Service (QoS) do not get met and user satisfaction drops. Traditional channel allocation and call admission control methods provide QoS guarantees to admitted sessions, allowing them to continue unimpeded during times of congestion, with the side effect that more user service requests are blocked when congestion occurs (Ahmed, 2005). Hou, Yang, and Papavassiliou (2002) study how to provide QoS guarantees to subscribers that provide both guarantees for admitted sessions as well as a low (less than 1%) overall call blocking probability. Hou et al. focused on making efficient use of the limited bandwidth and frequency spectrum of wireless networks while maximizing the social welfare of the network (i.e., admitting the most user sessions).

The key innovation provided by Hou, Yang, and Papavassiliou (2002) is the introduction of a dynamic pricing function to the CAC algorithm. Hou et al. prove that there is an optimal arrival rate of traffic for any network and that congestion and the associated call blocking occurs when the demand is greater than ability of the network to handle or, in other words, the actual arrival (i.e., demand) rate is greater than the optimal arrival rate for the network. Then, assuming users are price sensitive, the logic is that by raising the price dynamically when the arrivals exceed the optimum, enough users will choose not to enter the system and the arrival rate is kept below the optimum leading to a congestion free system. The system is shown in Figure 1. The pricing scheme, $H(t)$, is designed to control the user behavior such that the arrival rate, $\lambda_n(t)$, results in a to the P_{nb}, less than the required 1%.

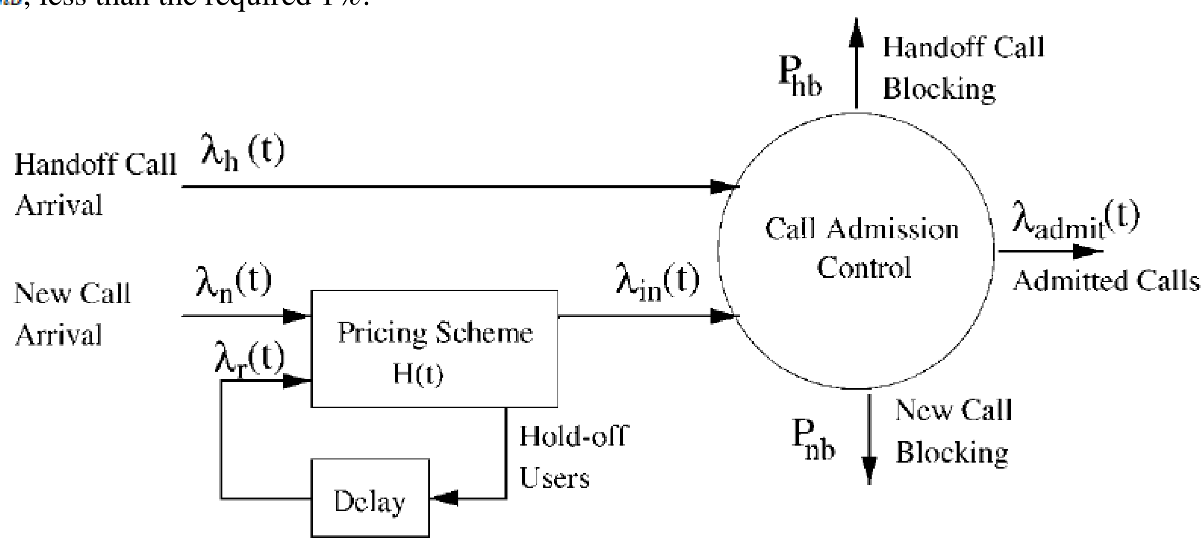

Figure 1. Dynamic Pricing with Call Admission Control (Hou et al. 2002)

Building on prior work (Hong & Rappaport, 1986; Hou, Yuguang, & Akansu, 2000; Kulavaratharasah & Aghvami, 1999; Lin, Mohan, & Noerpel, 1994), when the system parameters (total number of channels in the system, channels reserved for handoff calls, and the average holding times for new and handoff calls) are fixed, Hou et al. describe system performance in terms of the new call arrival rate. The authors then formulate the number of users admitted to the system solely as a function of the new call arrival rate and compute the resultant new and handoff call blocking probabilities, P_{nb} and P_{hb}, respectively, as functions of the new call arrival rate. Intuitively, the probability of blocking is low when the new call arrival rate is low and when the new call arrival rate is high, the probability an individual call is blocked is also high.

Hou et al. (2002) focus on the network ability to initiate a new call and maintain an in-progress call as the desired Quality of Service metric from the user perspective. The P_{hb}, as the Quality of Service metric which, as stated above, can be described as a function of the new call arrival rate.

In a practical application, a specific user utility function representing the user's satisfaction with the service is required. For their proof, rather than choosing a specific user utility function, the authors describe the nature of the utility function. The user satisfaction with the network service increases with a decrease of call blocking probability and vice versa. When the call blocking probability is low, user satisfaction decrease due to an increase in blocking probability is not significant, whereas when the call blocking probability is high, any increase yields a significant decrease in user satisfaction. The highest user satisfaction is achieved when the call blocking probability is zero. For all practical applications, there is a maximum probability at which the user satisfaction becomes zero.

Hou et al. (2002) view the cellular network as a universal community resource that should maximize the overall benefit of the community. For this reason, the authors' goal is to maximize the total user utility. That is, to maximize the total aggregate user utility (number of admitted users times the individual user utility above). Modeling the average number of admitted users and the call blocking functions as differentiable, monotonically increasing continuous functions of the arrival rate and the utility function of a single user as a differentiable, monotonically decreasing concave function of the call blocking probability, Hou et al. (2002) prove that there exists an optimal new call arrival rate that maximizes the total user utility (number of admitted users times the individual user utility). When the actual call arrival rate is less than the optimal rate, the admitted users get a better service quality whereas when the new call arrival rate is greater than the optimal rate, the quality degrades and may no longer be acceptable. This represents the congestion state.

The incorporation of a dynamic pricing function to the admission process by Hou et al. (2002) is used to control the user arrival rate. When the arrival rate is lower than the optimal rate, a base price is charged that is acceptable to all users. When the arrival rate is higher than the optimal rate, a higher dynamic peak hour price is determined based on an estimate of a user demand function such that the users that do not wish to pay this peak hour price will avoid using the network service, thereby reducing the actual new call arrival rate below the determined optimal rate, yielding a congestion free system.

Subsequent researchers have continued the work of Hou et al. through experimentation with the combination of dynamic pricing with different CAC schemes (Manaffar, Bakhshi, & Pilevari, 2008; Olivré, 2004) demonstrating the successfulness of the approach across these different CAC methods.

A major limitation of Hou, Yang, and Papavassiliou (2002) is a focus on a single class of service. In order to be widely applicable in today's third generation (3G) and fourth generation (4G) networks, support for multiple classes of service is required as indicated in subsequent research efforts that expand the dynamic pricing problem area (AbuAli et al., 2010; Al-Manthari, 2009; Al-Manthari et al., 2009; Chen, Ni, Qin, & Wei, 2010; Hou, 2003; Ozianyi, Ventura, & Golovins, 2008). Recognizing the need to support applications with differing QoS requirements, in his dissertation, Hou (2003), studies the problem of maximizing total user utility in a multi-class service network. While Hou expands his earlier theorem for optimal arrivals to support a multi-class service network, there are no practical experiments that validate the solution to the problem.

Al-Manthari (2009) and Al-Manthari, Nasser, Ali, et al. (2009) extend the research problem of Hou et al. (2002) for application to multi-class networks with the goal of both preventing congestion and increasing network utilization. The work also introduces a goal of fairness between service classes in utilization of the network. AbuAli et al. (2010) extend the concept of fairness in two directions: between classes and between connections within a class. Unlike the other multi-class service network studies, Chen, Ni, Qin et al. (2010) and Ozianyi, Ventura, and Golins (2008) relax the QoS guarantees per price for the user through the introduction of a congestion charge applicable to all users.

While, all of the multi-class service studies effectively use dynamic pricing to reduce congestion as expressed by call blocking to acceptable levels, they suffer from the same limitation. When the network resources are utilized to their most efficient, at some point, with increasing numbers of users, the price will become high enough to encourage the user to seek alternative operators.

Wenan, Ziaoli, Bing et al. (2010) demonstrate the effects of competing operators with heterogeneous networks in which each operator, seeking to maximize revenue employ dynamic pricing to prevent congestion, experiences that customers substitute similar services across operators as the prices of their preferred operator rise to reduce congestion.

Rather than using the heterogeneous network technologies in a competitive environment as in Wenan et al., operators are exploring how to utilize the heterogeneous technology in a cooperative manner to increase an individual operator’s available bandwidth.

Falowo, Zeadally, and Chan (2010) study the application of dynamic pricing to a joint CAC algorithm with the goal of improving the radio network efficiency by balancing the load on two different networks which has the side effect of helping the congestion problem, but without an explicit mechanism to prevent congestion, Falowo et al. fall short of providing a reduction in new or handoff call blocking probabilities.

None of the prior research addresses the congestion prevention problem in the current environment in which a wireless operator utilizes both broadband wireless access (e.g., LTE) and wireless local area network (i.e., WLAN) technologies. The research in this report addresses the problem of providing both guarantees for admitted sessions as well as preventing congestion with a low (less than 1%) overall call blocking probability in a cooperative heterogeneous multi-class service network while maximizing the social welfare of the network.

Barriers and Issues

Providing quality of service to users in a wireless cellular network, in general, is difficult for operators to achieve (Ahmed, 2005; Soldatos et al., 2005). For wired networks, the classic method is to over-provision the network. When bandwidth is insufficient in one part of the network, i.e., congestion is detected, the operator can simply increase the available bandwidth by adding another fiber to that part of the network and congestion is resolved.

In wireless networks, the spectrum available to provide bandwidth is limited by regulation and an increase is not possible or can be very cost-prohibitive. The other major influence is mobility. These two factors lead to less predictable availability and less bandwidth. CAC mechanisms to determine when traffic for a user is to be admitted to the network are identified to ensure that QoS guarantees offered to previously admitted users is maintained. An effective algorithm should provide accurate predictions in order for large number of users to make the optimal utilization of the limited resources. This algorithm, however, must also be easy enough to implement in a real-time system and often cannot due for a large problem size with complicated system parameters (Ahmed, 2005; Soldatos et al., 2005). The challenge becomes balancing this complexity with QoS guarantees and efficient resource utilization (Lima, Carvalho, & Freitas, 2007).

Additional challenges are presented with the rapid increase in content-rich applications each with distinct QoS requirements (Ahmed, 2005; Al-Manthari et al., 2011; Lima et al., 2007; Niyato & Hossain, 2005; Soldatos et al., 2005). Multiple service classes are required of the network in order to support multimedia applications – voice, video, audio, and data – extending the complexity required of admission control mechanisms controlling the limited network bandwidth determining which applications can be admitted to the network based on a prediction of the consumption by the application using the service. Ahmed (2005) and Al-Manthari, Nasser, and Hassanein (2011) note that this requires introduction of complex issues involving prioritization, fairness, and resource sharing.

For the evolving network, 4G technologies will be operating side-by-side with other wireless technologies, such as WLAN (Ahmed, 2005; Niyato & Hossain, 2005; Soldatos et al., 2005). This composite heterogeneous network provides additional opportunity for bandwidth expansion for offering QoS, with increased complexity of managing this new composite network.

Finally, pricing of services, in general, and the congestion pricing proposed in this paper, is complex and multifaceted. The operator must deal with a set of economic issues to provide incentives for network utilization and for accurate admission control (with or without dynamic pricing for congestion control) knowledge of user demand function and sensitivity to price is required (Al-Manthari et al., 2011).

The research in this report faced the same issues of multi-media applications with differing QoS requirements across a composite heterogeneous network, as well as pricing to resolve the congestion prevention problem. First, the network technologies were limited to LTE and WLAN. These are both technologies with well-understood interactions defined by 3GPP. Second, use of the simple CAC mechanism used by prior work of Hou (2002) reduced the complexity of computation by requiring only local cell-level information. Third, the set of service classes were derived and a limited set was used in the simulation. Finally, the pricing function was based on a standard demand model used in prior work in order to simplify the complexity of understanding user behavior in setting the price during times of congestion.

Dissertation Goal

The dynamic pricing with CAC algorithm introduced by Hou, Yang, and Papavassiliou (2002) prevents congestion in a single-class service network. Through simulation, Hou et al. demonstrate that the combined new call and handoff call blocking probability is reduced from 8% to less than 1% while maintaining the quality of service of previously admitted calls. Subsequent work extended this mechanism for use with the combination of dynamic pricing with different CAC schemes (Manaffar, Bakhshi, & Pilevari, 2008; Olivré, 2004); application to multi-class service networks (AbuAli et al., 2010; Al-Manthari, 2009; Al-Manthari et al., 2009; Ozianyi et al., 2008); and application to competitive heterogeneous multi-class service networks (Wenan et al., 2010).

The goal of this work was to extend the dynamic pricing with CAC algorithm mechanism for cooperative heterogeneous multi-class service networks in order to prevent congestion achieving a combined new and handoff connection blocking probability less than 1% as required by the International Telecommunications Union while maintaining the quality of service of previously admitted connections. In this work, the single-class service model was extended along two dimensions. First, a multi-class service model was defined and, second, a dual radio access technology model was defined. A dynamic pricing function was combined with the CAC function for this extended service model and, through simulation, demonstrated to achieve the desired 1% combined new and handoff connection blocking probability.

Relevance and Significance

Recent studies demonstrate that real-time entertainment is the dominant application on the Internet with P2P file sharing and web browsing still high in many areas (Sandvine, 2013). These three applications, in particular, have different requirements on the network for the delivery of data.

Even though the new technologies for broadband wireless access, such as LTE, support significantly higher data rates than their predecessors, researchers expect the increasing usage of content-rich bandwidth-intensive multimedia services will cause network congestion with these new technologies (Al-Manthari, 2009; Al-Manthari et al., 2009; Manaffar, Bakhshi, & Pilevari, 2008; Olivré, 2004; Ozianyi et al., 2008).

Many researchers have studied channel allocation and call admission control schemes that limit the number of users admitted into a network in order to reduce the amount of congestion in the network and deliver quality service to these users. These methods, however, are not able to guarantee an acceptable new and handoff call blocking probability to users (AbuAli et al., 2010; Al-Manthari, 2009; Al-Manthari et al., 2009; Al-Manthari et al., 2011; Hou, 2003; Olivré, 2004).

Several researchers have studied methods to reduce the blocking probabilities of both new and handoff calls with various alternative channel allocation and call admission control (CAC) schemes (Chang, Su, & Chiang, 1994; Hong & Rappaport, 1986; Hou et al., 2000; Kulavaratharasah & Aghvami, 1999; Lin et al., 1994). These schemes do not guarantee the acceptable 1% probability as defined by the International Telecommunication Union.

Unless some mechanism can be provided to shape the traffic inbound to the cellular system, it is not possible to prevent congestion of the system. Hou observed that the new call blocking probability increases with call arrival rate and that during the busy hour of the network none of the existing CAC schemes are able to guarantee an acceptable QoS.

Hou explains this in economic terms, that when there are no consequences, a user will be greedy in usage of the network which leads to the problem of demand for network resources exceeding available capacity.

Pricing of network services that influence user behavior has been used by operators to manage resources, enable provisioning of QoS for users, and improve system utilization (Courcoubetis & Weber, 2003). Courcoubetis and Weber note that pricing can be used to provide a control feedback function such that when prices are raised, demand is lowered and vice versa. When pricing incentives (and consequences) are offered, then users will behave more rational and police their own access to the network in order for the overall demand in an area to match the available bandwidth resources of the network in order to provide the most users with quality service (AbuAli et al., 2010; Al-Manthari, 2009; Al-Manthari et al., 2009; Courcoubetis & Weber, 2003; Gizelis & Vergados, 2011; Hou, 2003; Hou et al., 2002; Olivré, 2004; Ozianyi et al., 2008; Wenan et al., 2010).

Another method mobile operators have been utilizing to combat the surge of data network traffic is the availability of alternate wireless access technologies such as WLAN to offload the primary cellular system, such as LTE, when available. The integration of WLAN and cellular technologies has been a research topic with the earlier cellular technologies, GPRS and CDMA (Buddhikot et al., 2003; Salkintzis, Fors, & Pazhyannur, 2002), has been formalized in 3GPP standards (3GPP, 2009, 2010b) and is being deployed by operators such as KT Corporation in Korea (Choi et al., 2011).

Pricing as an incentive for users to offload traffic to the WLAN is also important (Falowo et al., 2010; Joseph, Manoj, & Murthy, 2004). The operators, however, are concerned that using the WLAN offload techniques will also cause congestion in this environment and are looking for an offload strategy that takes into account the congestion status of the WLAN as well as the congestion status of the LTE network (3GPP, 2011).

The work of Falowo et al. demonstrates the successfulness of integrating dynamic pricing into a joint channel admission control algorithm across heterogeneous networks for the purpose of balancing the load across the two network types which has the overall effect of better system utilization with a reduction in the new and handoff call blocking probabilities. The work, however, falls short of congestion prevention target offered by other work (AbuAli et al., 2010; Al-Manthari, 2009; Al-Manthari et al., 2009; Hou, 2003; Hou et al., 2002; Manaffar, Bakhshi, & Pilevari, 2008; Olivré, 2004).

The extension of the dynamic pricing with channel admission control documented in this report is a mechanism to fill the gap, successfully addressing congestion prevention across the combined heterogeneous LTE and WLAN networks.

Summary

In order to meet the increasing demands of mobile data users, a dynamic pricing function integrated with CAC function serves to achieve less than 1% blocking probability in a multi-class, dual radio technology network as documented in this report. In Chapter 2, a review of the literature in the use of dynamic pricing in mobile wireless networks is presented. In this review, the prior art is described with its limitations and how this prior art differs from the method in this report. Chapter 3 provides a generalized model for the dynamic pricing combined with CAC, describes the analytical model that defines the wireless network based on multiple service classes and dual radio access technologies introduced by this work and the experimental design validating this model through simulation. The results of the simulations demonstrating the successful resolution of the blocking problem are presented in Chapter 4. Final conclusions and recommendations in Chapter 5 suggest future work that can be explored to extend the method in this report.

Chapter 2

Review of the Literature

Dynamic pricing for congestion control in the CAC process, introduced by Hou, Yang, and Papavassiliou (2002), serves as the starting point for the literature review in a categorization of the work termed direct admission control. The direct admission control category focuses on the initial works involving a single-class service network (Hou et al., 2002; Manaffar, Bakhshi, & Pilevari, 2008; Olivré, 2004). This work serves as the basis for demonstration that congestion prevention in the form of a combined new and handoff call blocking probability less than 1% is achievable. A set of benchmark experiments are defined that subsequent works use as a basis for evaluation.

This literature review tracks the progression of the prior work from the initial single-class service networks to multi-class service networks. With multi-class service networks comes the additional problem of how to effectively allocate resources, not only among users as in the initial direct call admission control work, but across the multiple service classes available with evolved network technology. Two different types of allocation are reviewed: fixed allocation of resources among services classes (Al-Manthari, 2009; Al-Manthari et al., 2009; Hou, 2003) and dynamic allocation of resources among service classes (AbuAli et al., 2010).

An orthogonal extension is the introduction of congestion charges across all service classes (Chen et al., 2010; Ozianyi et al., 2008) leading to dynamic responses by users already admitted into the system. These extensions are analyzed for their effectiveness in solving the congestion problem while providing QoS guarantees to admitted users. The researchers in this area apply their research to a variety of different radio technologies including HSDPA (Al-Manthari, 2009; Al-Manthari et al., 2009), WiMAX (AbuAli et al., 2010; Chen et al., 2010), and WLAN (AbuAli et al., 2010) showing the ability for the model to be applied generally across network technologies.

Finally, current research showing the progression to heterogeneous wireless networks is presented (Falowo et al., 2010; Wenan et al., 2010). Heterogeneous networks with more than one radio technology are overlaid in a single cell. These networks may be owned by different operators and behave competitively or may be owned by the same operator and behave cooperatively. The competitive network research is presented in which the use of dynamic pricing is used to maximize the revenue of each of the competing operators as opposed to ensuring that congestion is prevented (Wenan et al., 2010). The cooperative network research presents a method of incorporating dynamic pricing in order to balance the load across the heterogeneous networks which does lead to network efficiency but does not lead to congestion prevention.

The work in this report fills the gap of dynamic pricing for congestion prevention in the cooperative heterogeneous network.

Direct Call Admission Control

Integration of dynamic pricing for congestion control into the CAC process, known as Direct Call Admission Control by Al-Manthari, Nasser, and Hassanein (Al-Manthari et al., 2011), have the common goals of efficient use of network resources with congestion control defined as reduction in call blocking probabilities. All the works described in this section utilize the basic dynamic pricing with CAC concept presented by Hou et al. (2002); employing a variety of different CAC functions from the literature in order to validate the common congestion prevention performance of the solution.

The difference between the proposed research identified in this paper and the prior work described below is the extension to parallel heterogeneous multi-class service networks. The fundamental concept introduced by the seminal work in this section – the combination of a dynamic pricing function with a CAC mechanism to prevent congestion – was utilized in the proposed research.

While not recent, the work of Hou et al. (2002) later expanded in Hou (2003) is seminal in the establishment of this research area and is included to provide the basis of comparison for the more recent work that follows. The main focus of the research of Hou et al. is to use dynamic pricing as an input to the congestion control problem to reduce the new call blocking probability for a network.

The first key contribution of Hou et al. (2002) is the establishment and proof of a theorem in which there exists an optimal new call arrival rate that maximizes the aggregate social utility (i.e., satisfaction) of the network service. The aggregate social utility is represented by the total user utility of the network, where user utility is defined as a function of new call and handoff call blocking probabilities.

Hou et al. (2002) uses the following methodology elaborated below:

1. Define system performance in terms of the new call arrival rate.
2. Define the Quality of Service metric as the weighted sum of the new and handoff call blocking probabilities.
3. Define the nature of the user utility function that describes individual user satisfaction with the defined QoS metric.
4. Model the average number of admitted users as a function of the arrival rate and define the aggregate user utility equal to the average number of admitted users multiplied by the individual user utility.
5. Maximize the aggregate user utility to determine the optimal arrival rate.

Starting with prior work (Hong & Rappaport, 1986; Hou et al., 2000; Kulavaratharasah & Aghvami, 1999; Lin et al., 1994), that is, when the system parameters (total number of channels in the system, channels reserved for handoff calls, and the average holding times for new and handoff calls) are fixed, Hou et al. describe system performance in terms of the new and handoff call arrival rate. Hou et al. go one step further demonstrating that the handoff call arrival rate can be determined as a function of the new call arrival rate.

With this as a starting point, the authors can formulate the number of users admitted to the system solely as a function of the new call arrival rate and compute the resultant new and handoff call blocking probabilities as a function of the new call arrival rate. Intuitively, this indicates that the probability of blocking is low when the new call arrival rate is low and when the new call arrival rate is high, the probability an individual call is blocked is also high.

Hou et al. (2002) focus on the network ability to initiate a new call and maintain an in-progress call as the desired Quality of Service metric from the user perspective. The authors use a weighted sum of the new and handoff call blocking probabilities as the Quality of Service metric which, as stated above, can be described as a function of the new call arrival rate.

In a practical application, a specific user utility function is required to express the user satisfaction with the call blocking probability. For their proof, rather than choosing a specific user utility function, the authors describe the nature of the utility function. The user satisfaction with the network service increases with a decrease of call blocking probability and vice versa. When the call blocking probability is low, user satisfaction decrease due to an increase in blocking probability is not significant, whereas when the call blocking probability is high, any increase yields a significant decrease in user satisfaction. The highest user satisfaction is achieved when the call blocking probability is zero. For all practical applications, there is a maximum probability at which the user satisfaction becomes zero.

Hou et al. (2002) view the cellular network as a universal community resource that should maximize the overall benefit of the community. For this reason, the authors are looking to maximize the total aggregate user utility, that is, to maximize the number of admitted users times the individual user utility above.

Modeling the average number of admitted users and the call blocking functions as differentiable, monotonically increasing continuous functions of the arrival rate and the utility function of a single user as a differentiable, monotonically decreasing concave function of the call blocking probability, Hou et al. (2002) prove that there exists an optimal new call arrival rate that maximizes the total user utility (number of admitted users times the individual user utility). When the actual call arrival rate is less than the optimal rate, the admitted users get a better service quality whereas when the new call arrival rate is greater than the optimal rate, the quality degrades and may no longer be acceptable. This represents the congestion state.

The second key contribution is the development of a model that incorporates a dynamic pricing module to the admission process as shown in Figure 1. When the arrival rate is lower than the optimal rate, a base price is charged that is acceptable to all users. When the arrival rate is higher than the optimal rate, a higher dynamic peak hour price is determined based on an estimate of a user demand function such that the users that do not wish to pay this peak hour price will avoid using the network service, thereby reducing the actual new call arrival rate below the determined optimal rate, yielding a congestion free system.

In Hou et al. (2002), the resources are priced according to the current traffic load based on a well-defined demand function (Fishburn & Odlyzko, 1998) that represents a user's willingness to pay for an offered service. In practical terms, Hou et al. propose that the user is informed of the price through cell broadcast in order to be able to choose whether to make a call or not.

Hou et al. incorporate a user response behavior into the model. The users that do not wish to pay the current price can choose to wait and make their call later. These retry users generate an additional traffic load on the system to be accounted for in the pricing function. The price influences user behavior to reduce the load on the system such that it is less than or equal to the optimal load.

As described, the dynamic pricing function incorporates the new call load plus the retry load. When the user accepts the price, the CAC block is invoked to determine if there are network resources available for the call. A simple guard channel CAC scheme in which a fixed number of channels are reserved for handoff calls is used to provide preference to handoff calls. When the number of available channels is less than the number reserved for handoff calls, a new call is blocked. A handoff call is blocked when there are no channels available in the system. These conditions are represented by the new call and handoff call blocking probabilities.

To demonstrate the effectiveness of the solution, Hou et al. (2002) perform simulations of a single cell with five different scenarios comparing conventional CAC without dynamic pricing to the new CAC with dynamic pricing system. Three different user behaviors are studied: retry; retry with loss; and holdoff retry. The first two apply to both the conventional CAC and pricing system with CAC. The third, holdoff retry, applies only to the pricing system with CAC. The experiments are summarized in Table 1. The experiment labeled PSwHR represents the specific model shown in Figure 1.

Table 1. Summary of Experiments in Hou et al. (2002)

Experiment	**Dynamic Pricing Included?**	**CAC**	**User Behavior**	**Description**
CSwR	No	Guard Channel	Retry	Blocked users retry after a wait (delay) time.
CSwRL	No	Guard Channel	Retry and Loss	1/3 blocked users exit system 2/3 blocked users retry after a

				wait (delay) time.
PSwHR	Yes	Guard Channel	Holdoff Retry	User that does not accept price waits for a while and retries. Users that accept price and are blocked by CAC, exit the system.
PSwR	Yes	Guard Channel	Retry	Hold-off and blocked users all retry after a wait (delay) time.
PSwRL	Yes	Guard Channel	Retry and Loss	1/3 of the holdoff users and 1/3 of blocked users exit 2/3 hold-off and 2/3 blocked users wait and retry.

Two conventional CAC scenarios without dynamic pricing for congestion are studied, one with full user retry after a wait time and a second with partial user retry with one third of the users leaving the system. Three dynamic pricing scenarios are studied that vary on the user response to a pricing increase. In the first, any user that is not willing to pay the higher price waits for a while and retries and any user that is willing to pay the higher price and is ultimately blocked due to channel unavailability leaves the system. The second dynamic pricing scenario mimics the first non-dynamic CAC scenario. Specifically, all users that hold-off and are blocked by CAC wait some time and retry. The third dynamic pricing scenario parallels the second non-dynamic CAC scenario. That is, one third of the users that hold-off and one third of the users blocked by CAC leave the system while the rest retry at a later time. For consistency, all five test scenarios utilize the same fixed guard channel scheme for handoff calls.

In order to utilize the model presented by Hou et al. (2002), the following characteristics of a cell must be predefined: cell capacity (i.e., number of channels); channels used for the service (e.g., one channel is dedicated for each call); call duration (e.g., exponential distribution with 240 second mean); cell dwell time (e.g., exponential distribution with 120 second mean); priority weighting of handoff calls vs new calls (e.g., two-to-one); and a user utility function. With these parameters defined, an optimal arrival rate that maximizes total user utility can be computed.

In simulations, Hou et al. defined the above parameters such that an optimal Poisson arrival rate was found to be 0.12 calls/sec. and tested the performance of the dynamic pricing algorithm with a Poisson process with rate varying from 0 to 0.15 over a 24 hour period with noon being the highest to determine the behavior during off-peak and peak (i.e., congested) periods.

In practical terms, Hou et al. (2002) note that in a real system, it is difficult to obtain the actual load at any given point in time. For this reason, they determine the average traffic load during a 5-minute period and use this to approximate the actual offered load in order to determine the price that is used for the next 5-minute period.

Hou et al. (2002) propose two kinds of utility functions: one representing an "inelastic" QoS requirement and the other representing an "elastic" QoS requirement. The utility function representing an inelastic requirement indicates a user that will not accept a service when its quality is below a specified threshold, such as new call blocking probability greater than 1% described by U_1, shown in Figure 2.

$$U_1 = \begin{cases} 1 - e^{30(P_b - 0.1)}, & when\ 0 \le P_b \le 0.01 \\ 0, & when\ P_b > 0.01 \end{cases}$$

An elastic, or soft, QoS requirement is represented by a utility function that decreases user satisfaction when the new call blocking probability exceeds 1% and approaches 0 as the probability approaches 0.1% as in U_2, shown in Figure 3.

$$U_2 = max\left(1 - e^{30(P_b - 0.1)}, 0\right)$$

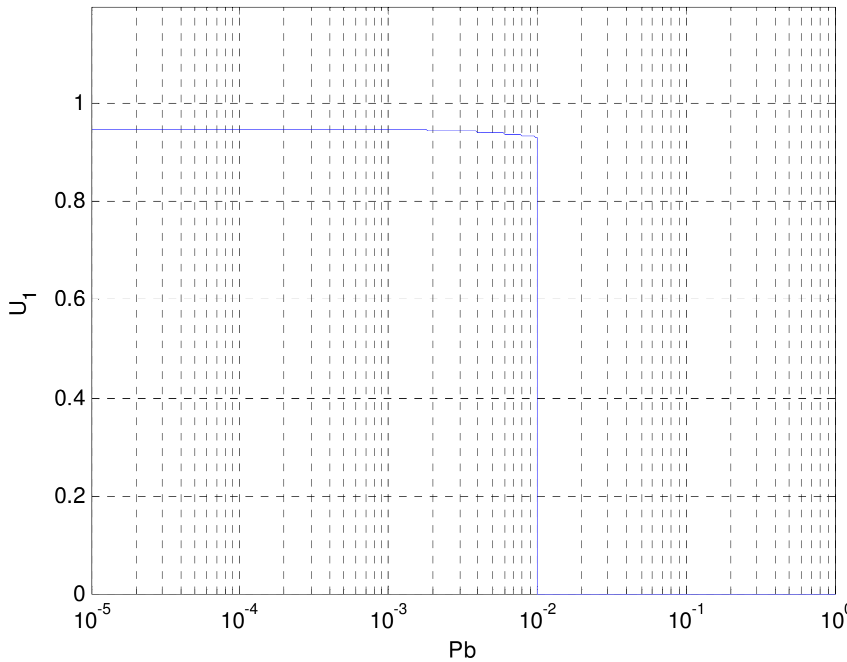

Figure 2. Inelastic QoS utility function

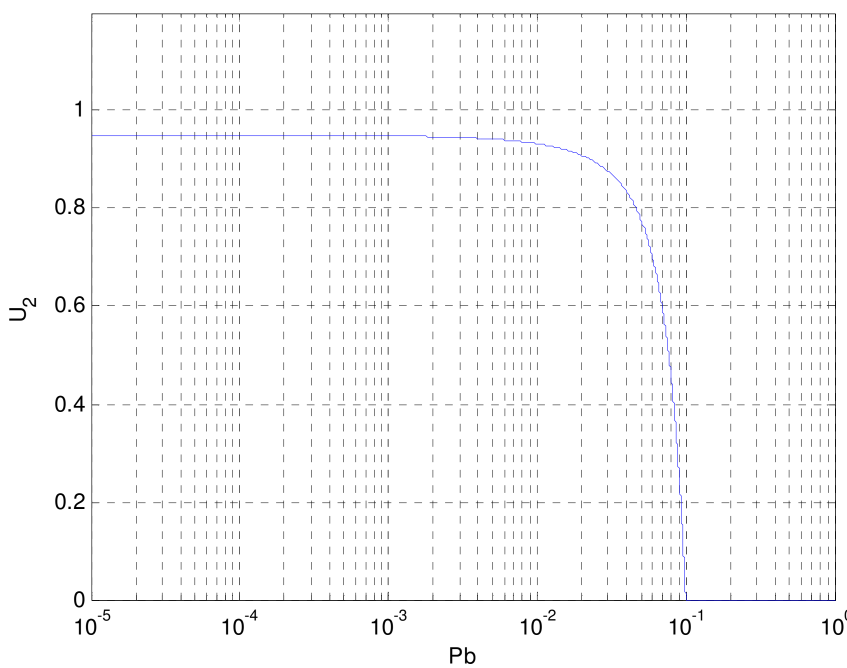

Figure 3. Elastic QoS utility function

For their experiments, Hou et al. (2002) use the inelastic QoS requirement offered by U_1 for the individual subscriber utility function.

In all three dynamic pricing scenarios, the results demonstrated that the solution prevented congestion, whereas the two conventional CAC scenarios could not. The authors also demonstrate that operator revenues increase as well even though fewer users are accepted into the system; they are paying a higher price during periods of congestion.

In their work, Hou et al. (2002) utilize a guard channel scheme for the CAC block. The separation of the CAC block from the pricing block allows for the incorporation of different CAC mechanisms with the dynamic pricing function.

Olivré (2004) studies the application of dynamic pricing to a variety of CAC methods including: the guard channel scheme used in Hou et al.; three queuing schemes in which new calls are queued, handoff calls are queued, and both types of calls are queued; channel borrowing; and dynamic and hybrid channel assignment. In the same manner, Manaffar et al. (2008) incorporates the finite priority queuing scheme of Chang, Su and Chiang (1994) as the CAC mechanism.

Olivré runs simulation experiments for GSM[1] voice calls using each of the CAC methods with and without the dynamic pricing component in order to determine the benefit of the dynamic pricing component. Olivré incorporates the same utility and pricing functions used in Hou et al. (2002). The work simulates a cellular system of 7 cells that wrap around to simulate a larger network as opposed to the single cell in Hou et al.

[1] Global System for Mobile Communications, known as a second generation mobile telecommunications transports speech over a fixed set of digital channels in an area known as a cell.

Table 2. Experiments Performed by Olivré (2004)

Experiment	CAC	User Behavior	Description
0 Guard Channel	No guard channel	Retry and Loss	Base experiment to determine the optimal arrival rate for the CAC system.
N Guard Channel	Guard channel	Retry and Loss	Repeat of CSwRL and PSwRL experiments of Hou et al. (2002) studying impact of various values for the number of guard channels.
Queuing of New calls	Queuing	Retry and Loss	
Queuing of Handoff calls	Queuing	Retry and Loss	
Queuing of Both	Queuing	Retry and Loss	
Channel Borrowing without Locking	Channel Borrowing	Retry and Loss	When a cell has no channel available, an attempt is made to borrow a channel from a neighboring cell.
Channel Borrowing with Locking	Channel Borrowing	Retry and Loss	When a cell has no channel available, an attempt is made to borrow a channel from a neighboring cell.
DCA	Dynamic Channel Allocation	Retry and Loss	A pool of channels is available for a set of cells to utilize and is dynamically allocated when on call arrival.
HCA	Hybrid Channel Allocation	Retry and Loss	Each cell has a fixed set of channels allocated with a shared pool of channels is available for dynamic allocation.

Olivré found that for GSM calls with all CAC variations, the introduction of the dynamic pricing function had a significant positive effect on the overall utility of the system, and almost always on the total revenue.

Of the different CAC methods, the best results were achieved with the queuing schemes both with and without the dynamic pricing component. Olivré also found that while the different CAC methods were good at increasing the utility, on their own, they did not increase the total revenue significantly. The introduction of dynamic pricing, however, succeeds in improving resource utilization for a resulting increase in revenue.

Manaffar et al. repeat the five experiments of Hou et al. using the finite priority queuing mechanism of Chang, Su, and Chiang (1994) with similar results and conclusions. The summary is presented in Table 3. The authors do not provide any significant details or insight into the contributions of the work.

Table 3. Experiments Performed by Manaffar et al. (2008)

Experiment	**Dynamic Pricing Included?**	**CAC**	**User Behavior**	**Description**
CSwR	No	Finite Priority Queuing	Retry	Blocked users retry after a wait (delay) time.
CSwRL	No	Finite Priority Queuing	Retry and Loss	1/3 blocked users exit system 2/3 blocked users retry after a wait (delay) time.
PSwHR	Yes	Finite Priority Queuing	Holdoff Retry	User that does not accept price waits for a while and retries. Users that accept price and a blocked by CAC, exit the system.
PSwR	Yes	Finite	Retry	Hold-off and blocked users all

		Priority Queuing		retry after a wait (delay) time.
PSwRL	Yes	Finite Priority Queuing	Retry and Loss	1/3 of the holdoff users and 1/3 of blocked users exit 2/3 hold-off and 2/3 blocked users wait and retry.

Multi-class Direct Call Admission Control with Fixed Class Allocation

Today's network is required to support a variety of applications, each requiring its own bandwidth and delay characteristics. The operator supports this by offering a set of service classes on a common shared network from which a particular application can choose. The work in this section is meant to control congestion while maximizing bandwidth utilization; reduce call blocking probability in a network with multiple service classes; providing fairness of network use among the different classes.

Two differences are identified between the work in this section and the research effort performed for this report. First, the work in this section is limited to a single multi-class service network and the proposed research effort will incorporate the use of multiple heterogeneous multi-class service networks. Second, the work in this section also attempts to maximize operator revenues in addition to controlling congestion. The research reported in this paper focused purely on congestion prevention and not on the revenue maximization problem.

Hou (2003) extends the theorem in Hou et al. (2002) to address multi-service networks in which each service has different bandwidth requirements. The author defines a set of service classes. Within each class is a set of service levels, each with a different amount of bandwidth required.

With a given defined traffic mix across all classes and levels, using the same methodology of their previous work, Hou proves that there exists an optimal new call arrival rate that maximizes the total utility. While Hou (2003) expands the optimal rate proof to multi-service environment, he does not include a full exploration of the dynamic pricing with integrated CAC model to the multi-service environment.

The application of dynamic pricing to multi-service networks is reflected in more current works that follow. (AbuAli et al., 2010; Al-Manthari, 2009; Al-Manthari et al., 2009; Ozianyi et al., 2008; Wenan et al., 2010)

Al-Manthari (2009) and Al-Manthari, Nasser, Ali, et al. (2009) extend the research problem of Hou et al. (2002) for application to multi-class networks with the goal of both preventing congestion and increasing network utilization. The work also introduces a goal of fairness between service classes in utilization of the network.

The authors define a system with a set of traffic classes. These classes each have different traffic characteristics, for example, one class may be defined for streaming data traffic, whereas another class defined for best effort data traffic, each of which impose different demands on the network for quality of service parameters. Within each traffic class, the system supports a set of services, each with its own bandwidth requirement. For example, the streaming data traffic class provides two services: one for audio and one for video with the bandwidth requirement for video being larger than the bandwidth requirement for audio. This service model is consistent with Hou (2003).

Based on the available bandwidth in the system at the current time, the authors define a scheme that computes the optimum number of connection requests for each service within each class to maximize usage of the available bandwidth and then uses dynamic pricing within each class to influence the arrival rate toward the optimum rate for the class for the next time window. Because the actual network condition is used as input to the determination of the optimal rate leading to congestion-free system, the authors claim that this results in the ability to remove the call blocking probability from the calculation.

Like, Hou (2003), this work utilizes a fixed traffic mix for all services across all classes and compute an optimal rate for each service. Unlike, Hou, the optimal connection rate for each service is determined that maximizes the use of the available bandwidth, removing the need to model individual user utility required to maximize total social welfare. On the other hand, this requires a computation in real-time of the optimal number of connection requests taking into account actual network conditions (i.e., available bandwidth) which is not required for the original methodology proposed by Hou et al. (2002).

Al-Manthari et al. (2009) evaluate the performance of their system using a discrete event simulation based on HSDPA.[2] Al-Manthari (2009) provides more detail about the simulation. The simulator is implemented in Java with the objective function that derives the optimal rate per service modeled as an integer linear programming problem that can be solved using *lp-solve*[3]. The author models a single cell independent of the rest of the system using a guard channel scheme for handoff calls and focuses the analysis on new calls in a single cell.

For this study, the authors use two traffic classes with three different services. Two streaming services are offered: audio (64 kbps) and video (384 kbps) and one best-effort service: FTP (128 kbps) with a traffic mix of ¼, ¼, and ½ for each of the services, respectively. Since the audio and video services are in the same class, this gives equal share to each service class. As done in Hou et al. (2002) and Olivré (2004), the study utilizes an actual arrival rate that varies over a 24-hour period with the highest rates in the middle of the day. This study utilizes a chart based on actual observations rather than the

---

[2] High-Speed Downlink Packet Access is an enhanced third generation system that allows for a variety of downlink speeds for data applications.

[3] Available from http://lpsolve.sourceforge.net/5.5/.

smooth curve used by previous works. As in all the prior work discussed, the connection arrival rates are modeled as a Poisson process with the mean total arrival rate as per the 24-hour graph. The duration of a connection is modeled as an exponential distribution with a mean value of 30 seconds. The authors adopt a pedestrian mobility model with a speed of 3km/hr.

Four performance metrics are studied: percentage of overall bandwidth utilization; connection blocking probability; percentage of bandwidth share by service class; maximum revenue earned during the simulation time. The authors demonstrate that their CAC-based dynamic pricing method has a 0% connection blocking at all times including those periods when the conventional CAC scheme experiences connection blocking approaching 20%. The result of zero is explained due to the fact that an optimum arrival rate for each class is determined based on the actual bandwidth available as opposed to a pre-determined optimum rate for the system as in Hou et al. (2002).

Multi-class Direct Call Admission Control with Dynamic Class Allocation

The research problem in AbuAli et al. (2010) is to control congestion while providing QoS guarantees for a multi-class service network in which there is both fairness between the classes and fairness among the individual connections within the class. The difference between the work in this section and the proposed research work is the support of multiple heterogeneous networks. The research performed for this report also maintained the fixed class allocation of previous work.

As discussed below, the algorithms presented divide the resources within a single multi-class network. This research work addressed the utilization of a combined pool of resources from multiple heterogeneous networks focusing specifically on congestion prevention and did not include resource management and scheduling algorithms as in the work in this section.

AbuAli et al. (2010) extend the concept of fairness in two directions: between classes and between connections within a class. The authors propose a real-time scheduling algorithm that incorporates two non-cooperative game-theoretic pricing algorithms. The first computes the optimum number of time slots available for each class of service in a multi-class network. The second computes the optimum number of time slots available for each connection within the class. These two algorithms for determining the optimal allocations replace the pre-defined ratios used in prior work (Al-Manthari, 2009; Al-Manthari, Nasser, Ali, & Hassanein, 2008; Al-Manthari et al., 2009; Hou, 2003; Olivré, 2004). Both computations are formed as pure Nash equilibria which have been demonstrated to be computed in polynomial time (Fabrikant, Papadimitriou, & Talwar, 2004).

As opposed to the CAC schemes studied in this paper, the algorithm controls congestion through reduction of the average throughput of the served connections while only providing QoS guarantees on the minimum data rate requested for a connection. Through simulation, the authors demonstrate that the scheme was able to achieve fairness and satisfy connection-level QoS for all connections in the system.

AbuAli et al. evaluate their method using a single cell of a WiMAX network (IEEE, 2009) with four classes: bi-directional voice streaming (i.e., VoIP); video streaming; FTP; and background best effort traffic. Four performance metrics are evaluated: average throughput; average queuing delay; packet loss; and fairness. The experiments are run with 9 different loading mixes of the four classes of traffic. The scheme was shown to control congestion while providing fairness and QoS guarantees at the class level. The scheme was also shown to be able to meet the QoS requirements with fairness at the connection level.

Beyond the evaluation simulations, the authors have implemented the schemes in a municipal WLAN test bed in which they plan to study the scheme further as a mechanism to prevent congestion as well as to study different types of applications (real-time and non-real-time).

Multi-class Direct Call Admission Control with Congestion Charge

Unlike the other prior works, those in this section affect the price charged to users already admitted to the network, relaxing the QoS guarantees per price for the user. The research work performed focused on using dynamic pricing for congestion prevention associated with new connections only and continued to provide QoS guarantees per price in the heterogeneous multi-class combined network.

Chen, Ni, Qin et al. (2010) extends the dynamic pricing CAC scheme for a multi-class network of Al-Manthari, Nasser, Abu Ali, and Hassanein (2008) with a network congestion control component that adds a congestion charge to all connections in the network when congestion occurs. Al-Manthari et al. (2008) presents essentially the same model as their later work (Al-Manthari et al., 2009) using a few simplifications.

With this added congestion charge, Chen et al. relaxes the QoS guarantee constraint of admitted connections in favor of increasing the overall utilization of the network. The impact of congestion is spread to all users, not just the newly arriving ones. Chen et al. does this by introducing a network congestion charge into the price for ongoing sessions in addition to dynamic pricing for the newly arrived calls.

When the network congestion charge is added or changed, users in the system have the opportunity to quit earlier than expected because the price is too high; give up some of their allocated resources in order to maintain the prior price; or keep the allocated resources, but pay the higher price. The behavior of users in response to a price increase for ongoing connections, therefore, could yield additional bandwidth resources that need to be factored into the optimization.

As in prior scenario of Al-Manthari et al. (2008), the scheme determines the optimum arrival of new services to maximize the network utilization based on the available resources taking into account the behavior of user responses to an increase in the network congestion charge in addition to the behavior of users arriving in the system.

Through simulation, the authors demonstrate better network utilization and increased revenue for the operator using their scheme over conventional CAC. The study compares the results of their scheme with the bandwidth management scheme of Al-Manthari et al. (2008) demonstrating that in spite of the increased network congestion cost, the utilization is higher and the resulting revenues are higher.

Chen et al. (2010) evaluate the performance of their scheme using a discrete event simulation of WiMAX (IEEE, 2009) using a single cell as in Al-Manthari et al. (2008). For their study, the authors use only two service classes: bi-directional streaming voice (64 kbps) and video on demand (128 kbps). Their study uses a realistic 24-hour model for arrival rates for a Poisson process which is very similar to that used for Al-Manthari et al. (2009).

The authors perform three experiments: conventional CAC scheme with fixed prices of 0.4 and 0.2 cents per unit bandwidth for the two classes; bandwidth management scheme of Al-Manthari et al. (2008); and the proposed dynamic pricing scheme with network congestion charge. The results demonstrated increased network utilization over the dynamic pricing scheme of Al-Manthari et al. The authors confirm the results demonstrated by Al-Manthari et al. for zero blocking probability during congestion and show that their scheme which introduces the network congestion charge has a small connection blocking probability (up to 6%) relative to the significantly larger blocking probability (up to 19%) of the conventional CAC. The authors demonstrate higher network utilization and revenues with the tradeoff of the introduction of a small connection blocking probability.

Ozianyi, Ventura, and Golins (2008) introduce a concept in which dynamic pricing is combined with different levels of QoS guarantee in order to accommodate congestion control along with user sensitivity to price. One key concept is that some users are less sensitive to price and when the price is raised for all users in order to manage congestion control; these users will accept the higher price in order to maintain guaranteed QoS. A second set of users will accept a downgraded service in order to maintain a consistent price. The third set of users pay a flat rate that is not usage-based and receive little to no QoS guarantees for traffic.

This separation is similar to the concept introduced in Chen et al. (2010) in which three types of users of a given class were identified: those that would accept a higher usage price to maintain the current QoS of their connection; those that would give up some resources (i.e., lower QoS) to maintain their price; and those that would refuse to pay the higher price and leave the system.

Ozianyi et al. have introduced another option, that of the flat-rate user that is satisfied if their traffic gets through eventually during time of congestion. Rather than having to deal with the user behavior in response to a price change in real-time as required by Chen et al., Ozianyi et al. codifies the user response in the service class such that the user requests the service class in advance telling the network what his behavior will be during times of congestion.

The second key concept is to allow resource sharing between profiles (Ozianyi et al., 2008). During times of congestion, the resources allocated to users of lower levels can be borrowed to maintain the QoS requirements of users at the higher levels. This resource sharing can be performed as long as the minimum guarantee of the lending class is still met, providing similar constraints as in AbuAli et al. (2010). In other studies, the bandwidth allocation between service classes is set to an operator-defined ratio (Al-Manthari, 2009; Al-Manthari et al., 2009; Hou, 2003; Olivré, 2004). Through emulation, the authors demonstrate increased number of connections in higher classes of service when resource sharing is enabled, leading to greater operator revenues.

Dynamic Pricing in Heterogeneous Network Environment

In order to reflect the current network environment, it is important to address the presence of multiple heterogeneous networks covering the geographic region. In some cases, the heterogeneous networks are provided by competing operators (Wenan et al., 2010). In others, the heterogeneous networks are provided by the same operator (Falowo et al., 2010). In both cases, the application of dynamic pricing may encourage users to utilize the services of one access network or the other.

The research work in this report was like Wenan, Ziaoli, Bing et al. (2010) and Falowo, Zeadally, and Chan (2010) in that the cell model incorporates multiple radio access technologies covering the same geographic region. Unlike Wenan et al. and similar to Falowo et al., the radio access technology networks were operated by a single network provider allowing for joint admission control. The research work differed from Falowo et al. in that the goal of the work used dynamic pricing to explicitly prevent congestion and, as in Wenan et al, users could actively choose to substitute the services of one radio access network for another when the price, set dynamically, for doing so was attractive.

Wenan et al. (2010) introduce competition through heterogeneous network access technologies. In a given geographic area, a user may be simultaneously covered by three different network access types such as WLAN, WiMAX, and UMTS. The authors introduce the concept of QoS sacrifice in which the network automatically downgrades users in order to admit new users to the network that will generate more revenue for the operator. Even though the operator pays a penalty fee for the QoS sacrifice and potentially loses the customer in that moment to one of the competing networks in that area, the admission of higher revenue users offset the penalties.

Wenan et al. (2010) pose the pricing problem as a competitive game between the operators serving the area. The results of the game indicate the prices that should be charged by each operator to achieve maximum revenue. Two different results to the game are studied. When all operators announce their price simultaneously, they will tend to be lower than when one operator is in a leading position and announces the price first and the other operators are forced to follow, resulting in different revenue profile for the operators.

Wenan et al. (2010) demonstrate through simulation that when the dominant operator in a region begins to experience congestion and raise prices, the bandwidth demand for the competitive networks increase to accommodate the users that switch networks. Since at the same time the operator is raising prices, it is also selectively admitting high profit service connections and downgrading less profitable service connections, the operator still experiences an increase in revenue.

The key component of Wenan et al. (2010) applicable to the proposed research is the substitutability of service classes across technologies. The ability of the user to decide to substitute a service class available on WLAN, for example, in place of LTE when the price of the alternate service class has a cheaper price than the one currently being utilized is important for the non-competitive scenario to be studied. Wenan at al. incorporate a substitutability matrix that indicates the degree to which specific service classes of one technology are swapped by a user. This matrix is an input into the pricing model used by Wenan et al.

For the research study, the substitutability matrix was reflected in the user behavior in response to the offered dynamic price. It was also considered when a user actively decided to substitute an alternate class for the current class without a provoking event. In Wenan et al., all users experience the impacts of congestion due to combined effects of QoS sacrifice and price increase for existing connections, thus triggering use of alternate service classes. In the performed study, the existing connection was not impacted, but the incentive offered on the new network may have been sufficient for the user to trigger a substitution.

Falowo, Zeadally, and Chan (2010) study the problem of resource utilization across collocated heterogeneous wireless networks in which an operator deploys multiple radio access technologies to provide ubiquitous coverage and high QoS. The authors define a joint call admission control mechanism that not only determines if a call can be admitted, but also must determine the radio access technology to utilize for session. The authors build on prior work that determines that an algorithm that load balances among multiple access networks leads to better radio network utilization. The authors incorporate user preferences for radio access technology, which could lead to an unbalanced network due to user favoring a specific technology, with a dynamic pricing algorithm that encourages equal load across the access networks. Like the works for homogeneous wireless networks above, Falowo et al. provide QoS guarantees for admitted calls (although not the price during handover) and priority to handoff calls before new calls.

The algorithm employed by Falowo et al. does not provide direct correlation to a subscriber's willingness to pay as in Hou et al. (2002), but one of the preferences input by the user is sensitivity to price. The actual arrival rate for a particular service class for a specific radio access technology is compared against an ideal arrival rate that balances the load for the service class across the access networks. When the ideal rate on one network is lower than the actual, the price associated with a unit of bandwidth on that network is incremented. When the next user wishes to enter the system, his preferences are assessed and the new price is used in the evaluation of which network to use. Thus, the price is used to control access to a network in order to balance the load across the networks. The price is repeatedly increased until the ideal rate is achieved.

Each service class offered by the operator supports a fixed set of bandwidth allocations such that when a call is admitted to a particular network, the maximum bandwidth supportable by the radio technology in the requested class is allocated. With this method, the user isn't necessarily allocated the amount requested; the user application must adapt to the committed bandwidth based on the radio technology of the network selected. For handoff calls, a certain bandwidth is reserved in order to give priority to handoff calls.

Falowo et al. analyze a single cell with three co-located networks, each utilizing a different radio access technology, using a markov decision model in MATLAB. A standard Poisson arrival process with mean duration is used as in the other work studied. The authors demonstrate that over time, the dynamic pricing leads to a balanced network as well as noticeable reductions in connection blocking probabilities for both new and handoff calls. When the user preferences indicate that the user has high price sensitivity, the dynamic pricing algorithm has a better effect on the load balancing then when the users have low price sensitivity. While the authors demonstrate that the call blocking probabilities drop, there is no attempt to prevent the congestion directly. There is an indirect effect on congestion due to the balancing of the networks.

Summary

Hou et al. (2002) serves as the seminal work in the application of dynamic pricing to CAC in order to prevent congestion. Hou et al. proved that a wireless network system has an optimal arrival rate that maximizes the social utility of the wireless network. The incorporation of a dynamic pricing function based on user willing to pay to the CAC function, provides an economic method of influencing user behavior such that the actual arrival rate approaches the optimal rate and prevent congestion in the form of a new call blocking probability less than 1%. The applicability of this work to today's wireless network is limited due to the evolution of wireless technology beyond the author's time of investigation. Olivré (2004) and Manaffar, Bakhshi, and Pilevari (2008) repeat the experiments performed by Hou et al. substituting different CAC algorithms, demonstrating the effectiveness of the dynamic pricing with CAC methodology to maintain the new call blocking probability less than 1%.

The wireless network has evolved to support multiple service classes. Hou (2003) began the extension of the prior work to the multi-class service network with the definition of service classes each with multiple levels of bandwidth requirements. With a fixed traffic mix across all classes and levels, using their previous work, Hou proves that there exists an optimal new call arrival rate that maximizes the total utility. This concept is augmented by Al-Manthari et al. (2009) and Al-Manthari (2009) who incorporate the available bandwidth into the computation of the optimal arrival rate requiring a complex calculation in real-time during the CAC process. AbuAli et al. (2010) presents an algorithm for determining fairness between classes as well as between connections within a service class rather than relying on the fixed class allocation of the prior work. Using a game-theoretic model, AbuAli et al. computes an optimal allocation of bandwidth between classes and a second computation for use within each class of service. The algorithm controls congestion through reduction of the average throughput of the served connections while only providing QoS guarantees on the minimum data rate requested for a connection. These works increase the amount of computation required in order to prevent the congestion in multi-class service networks.

As an alternative, Chen et al. (2010) and Ozianyi et al. (2008) introduce a congestion charge to previously admitted connections. As a result, the user connection may end up with a higher than originally committed cost to achieve the QoS requested. The users may accept the congestion charge or change to a lower class of service. Ozianyi et al. also introduce the concept of resource sharing between profiles in order to promote usage in higher classes of service. These methods lead to higher network utilization and revenues with the tradeoff of the introduction of a small connection blocking probability. These methods do not solve the problem of maintaining the new call blocking probability less than 1%.

Finally, the concept of heterogeneous wireless networks, as being deployed currently, is studied. Wenan et al. (2010) and Falowo et al. (2010) both address geographic regions in which multiple radio access technologies overlap allowing for additional opportunities for congestion prevention. In a competitive network scenario, Wenan et al. apply dynamic pricing in order to maximize operator revenues in times of congestion. The authors demonstrate the effects of a user substituting one network service class with another as the price rises showing how an operator maximizes revenue in this competitive environment. Falowo et al., on the other hand, demonstrates a cooperative environment in which the different radio access technologies are deployed by the same operator. In this scenario, the goal of the work is to balance the utilization of two different networks. Both methods maximize utilization of the networks; however, neither achieves the new call blocking probability less than 1%.

The research work documented in this report started with the multi-class service model of Hou (2003) and added a third dimension for cooperative wireless networks, as in Falowo et al. (2010) with the incorporation of the substitutability matrix of Wenan et al. (2010) to address user behavior for switching between equivalent service classes to obtain a better price. With these extensions, the inclusion of a dynamic pricing function with CAC successfully prevented congestion achieving a combined new and handoff connection blocking probability less than 1% as required by the International Telecommunications Union while maintaining the quality of service of previously admitted connections.

Chapter 3

Methodology

The completed research followed an experimental design. The research studied whether the introduction of a dynamic pricing function into the connection admission control (CAC) process for a single multi-class service network through cooperative deployment of heterogeneous radio technology consisting of LTE and WLAN could be used to prevent congestion from occurring and reduce the overall call blocking probability for a user below the 1% target recommended by the International Telecommunication Union. The experimental design detailed in this chapter consisted, first, of an analytical model of the CAC algorithm with dynamic pricing in a heterogeneous environment. The analytical model was subsequently validated through simulation (Jain, 1991).

The remainder of this section presents the approach used for the work. First, a generalized model of dynamic pricing with call admission control derived from prior work is presented identifying those areas in which the research modified the model. Second is discussion of the analytical model that was developed representing the model. Third, the experimental design for the simulation phase of the work is described. The section concludes with discussion of the interpretation of the data collected from the simulation.

Generalization of Dynamic Pricing with Call Admission Control

The work using dynamic pricing for congestion prevention that builds upon Hou et al. (2002) followed the general system for a given cell presented in Figure 2. The first premise is that a lateral handoff call from one cell to a neighboring cell of the same radio technology was offered at the previously agreed price. Therefore, the incoming arrival rate for handoff calls was not subject to a dynamic pricing function and is shown direct into the CAC function. New calls, on the other hand, were subject to dynamic pricing, as depicted by the incoming arrival rate being subject to a pricing function. If the price was accepted, the connection was admitted based on available resources by a channel allocation and call admission control function. From either the pricing or CAC functions, the users that were not admitted to the system then had some response behavior that was incorporated into the system.

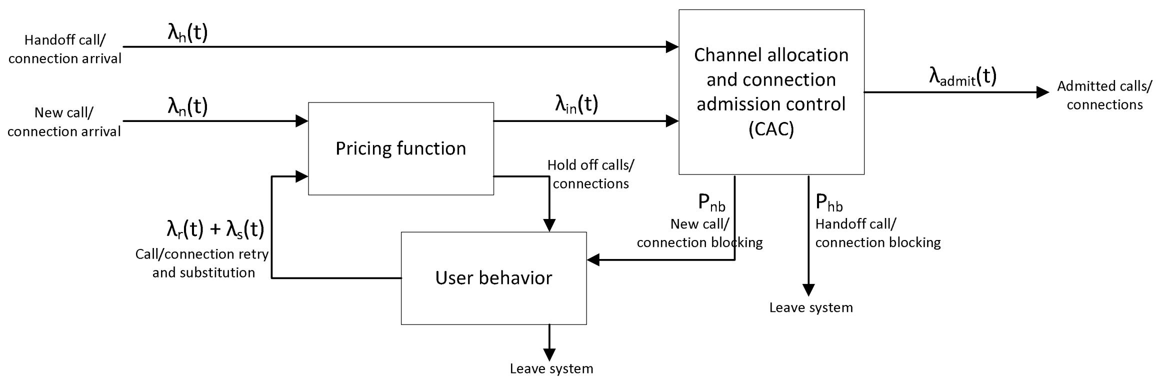

Figure 4. General Dynamic Pricing with CAC

Three components are depicted in the generalized model. Two are functional components of the system: pricing function and CAC function. The third component is a representation of user behavior that was incorporated in the model and associated experiments. The separation of CAC from pricing function was important to allow for the combination of any CAC method with a dynamic pricing function as demonstrated by Olivré (2004) and Manaffar et al. (2008). Since the selection of CAC method does not affect the outcome of the system, the guard channel CAC method from Hou et al. (2002) was used for the completed research. The guard channel CAC method consisted of reserving a number of channels for handoff calls only with the remaining channels available for both new and handoff calls. This reservation gave the handoff calls a higher priority over new calls.

A variety of pricing functions have been studied, but the main premise was that the pricing function is chosen based on user willingness to pay for the service with the goal of setting the price such that the arrival rate is reduced below the optimum arrival rate for the CAC system employed. In Hou et al. (2002), price is a function of the current traffic load consisting of new call requests plus retry requests from blocked users. This work extended the pricing function to incorporate the traffic load from substitution requests from users as used in Wenan et al. (2010) and Falowo et al. (2010). The other component of the price function in Hou et al. is a demand function representing the user's willingness to pay for the service. This work utilized the same demand function.

In response to the dynamically set price, new call blocking, or handoff call blocking, the user took some action. Table 4 identifies the set of user actions derived from the reviewed literature. Experiments including all of these user actions were incorporated into the work.

Table 4. User Actions in Response to Pricing or CAC Blocking Events

User Behavior	Description	Sources
Retry	Blocked users retried after a wait (delay) time.	Hou et al. (2002); Manaffar et al. (2008)
Retry and Loss	1/3 blocked users exited system 2/3 blocked users retried after a wait (delay) time.	Hou et al. (2002); Manaffar et al. (2008)
Holdoff Retry	User that did not accept price, waited for a while and retried. Users that accepted price and blocked by CAC, exited the system.	Hou et al. (2002); Manaffar et al. (2008)
Retry	Hold-off and blocked users all retried after a wait (delay) time.	Hou et al. (2002); Manaffar et al. (2008)

Retry and Loss	1/3 of the holdoff users and 1/3 of blocked users exited 2/3 hold-off and 2/3 blocked users waited and retried.	Hou et al. (2002); Olivré (2004); Manaffar et al. (2008)
Substitute and Retry	User that did not accept price substituted service and retried.	Wenan et al. (2010); Falowo et al. (2010)
Substitute and Handoff	User handed off to an alternate service with the possibility of a lower price.	Wenan et al. (2010); Falowo et al. (2010)

Analytical Model Design

The key components of the analytical model consisted of the definition of the service model, system parameters, and the pricing function. The derivation of the multi-class service model is presented first followed by the set of system parameters required for the model, and finally the development of the pricing function based on these parameters.

The service model for the research incorporated the three dimensions of application type, bandwidth level, and access technology. The analytical model of Hou (2003) addressed the two dimensions of application type and bandwidth level through the concept of a service class that identified the combination of application type and bandwidth level. Using the same technique, this work incorporated the access network into the definition of service class with each triple of (application type, bandwidth level, access network) represented by a unique service class. The paragraphs that follow describe the derivation of the service class model using Hou (2003), Falowo et al. (2010) and Wenan et al. (2010).

Hou (2003) defined a system model that supports $T + 1$ different types of applications, each with individual QoS and bandwidth requirements. The special case for voice using fixed channel size as used in Hou et al. (2002) was defined as type 0. Types 1 to T corresponded to T different types of data applications. Hou went on to define L_i, the number of different bandwidth levels available for each call of type-i application. This two-dimensional model served as a basis for the system model used in this research.

Falowo et al. (2010) and Wenan et al. (2010) introduced the concept of J radio access technologies with collocated cells. Falowo et al. supported multiple levels of bandwidth requirements for each type of application, but the different levels were associated with the different radio access technologies and were not selectable by the user as in Hou (2003). Wenan et al. supported continuous bandwidth utilization independent of the technology; as a result, the bandwidth level used by the application was available in either radio access technology. This research work extended the multiservice model for $T+1$ types of applications with L_i bandwidth levels of Hou to the third dimension of J radio access technologies of Falowo et al. maintaining the multiple bandwidth levels of Hou across all technologies as would be possible in Wenan et al.

Hou (2003), then, introduced the concept of a service class *k* which identified the combination of application type *t* and bandwidth level *i* resulting in $K+1=1+\sum_{i=1}^{T} L_i$ distinct service classes. The 1 is used to represent the type 0 voice service. Using the same technique, this research work incorporated the radio access technology *j* into the definition of service class *k* resulting in $K+1=1+J\cdot\sum_{i=1}^{T} L_i$ distinct service classes. Again the 1 was used to represent the type 0 voice service offered only on the primary mobile wireless technology.

With this service class structure, there were multiple service classes with equivalent application type and bandwidth level. Wenan et al. (2010) introduced a service substitutability matrix representing the user willingness to substitute the services of one access network with the services of another. In this work, the service substitutability matrix was used to represent equivalent classes that could be substituted for each other. This was used for modeling the user actions involving substitution described in Table 4.

Structuring the model as a set of service classes allowed for continued use of the model in Hou (2003). The system parameters required for the methodology in Hou consisted of the exponential distribution of the call duration times for new and handoff as well as the cell dwell time. The primary QoS parameter of study was the composite blocking probability which Hou defined as the weighted sum of the individual new call blocking probability, P_k^{nb}, and handoff call blocking probability, P_k^{hb}, $P_k = \alpha P_k^{nb} + \beta P_k^{hb}, k = 0, 1, \dots, K$, for each class of service k. The weighting factors α and β represented the relative cost or penalty associated with blocking the new or handoff call. In this study, it was twice as costly to block a handoff call ($\beta = \frac{2}{3}$) as it is to block a new call ($\alpha = \frac{1}{3}$). The individual user utility for a user of class k service, U_k, was a function of the composite blocking probability. If the number of admitted users of class k was represented by N_k, then the total user utility U, was defined by $U = \sum_{k=0}^{K} N_k U_k$. Hou defined a traffic mix as $\Gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)$ where $\sum_{k=0}^{K} \gamma_k = 1$. γ_i represented the probability that a given arrival was of class k. Hou proved that for a given traffic mix and utility function, there existed an optimal call arrival rate that maximized the total user utility. Hou demonstrated that this optimal call arrival rate could be calculated using a $K + 1$ dimensional Markov Chain model.

The system function embodied in the pricing component of Hou (2003) represented the percentage of users willing to pay the current price, commonly known as a demand function. The demand function for a single service class utilized by Hou was represented as $D[p(t)] = e^{-\left(\frac{p(t)}{p_0}-1\right)^2}, p(t) \geq p_0$ where p_0 was the normal price and $p(t)$ was the dynamic price computed to yield the optimal arrival rate at time t. Hou derived the pricing function based on the arrival rates for new calls, λ_n, and retry calls, λ_r, in comparison to the optimal new call arrival rate for the system, λ_n^*,$p(t) = D^{-1}\left(\min\left(\frac{\lambda_n^*}{\lambda_n(t) + \lambda_r(t)}, 1\right)\right)$. This research work applied this same demand function to determine the price for each service class in the model.

Since the function applied in this work introduced an additional arrival flow associated with equivalent service class substitution as shown in Figure 4, the pricing function was extended with the rate of substitute users, $\lambda_s(t)$,

$$p(t) = D^{-1}\left(\min\left(\frac{\lambda_n^*}{\lambda_n(t) + \lambda_r(t) + \lambda_s(t)}, 1\right)\right).$$

Experimental Design

The three components each represented a factor in the experimental design. The primary factor explored was the incorporation of the dynamic pricing component in the CAC algorithm. The secondary factor introduced was the selection of the CAC component itself. Prior work of Olivré (2004) and Manaffar et al. (2008) demonstrated that the specific CAC method chosen does not affect the outcome, thus, this was ruled out as an experiment factor.

A factorial design was used for the experiment with the two remaining factors (Jain, 1991). The first factor represented the inclusion of the dynamic pricing function (yes/no). The second factor represented the user behaviors listed in Table 4. Some of the user behavior actions, such as the user not accepting current price and waiting to retry, do not apply when the price is not dynamic, so these actions were not considered for the experiments in which the dynamic pricing component was not included. The full set of experiments performed is summarized in Table 5.

Table 5. Summary of Experiments Performed

Experiment	Dynamic Pricing Included?	User Behavior	Description
CSwR (Figure 5)	No	Retry	Blocked users retried after a wait (delay) time.

CSwRL (Figure 6)	No	Retry and Loss	1/3 blocked users exited system 2/3 blocked users retried after a wait (delay) time.
CSwSR (Figure 7)	No	Substitute and retry	Blocked users substituted a service and retried.
CSwSH (Figure 8)	No	Substitute and Handoff	Admitted user handed off to the alternate service class with the possibility of a lower price.
PSwHR (Figure 9)	Yes	Holdoff Retry	User that did not accept price, waited for a while and retried. Users that accepted price and were blocked by CAC, exited the system.
PSwR (Figure 10)	Yes	Retry	Hold-off and blocked users all retried after a wait (delay) time.
PSwRL (Figure 11)	Yes	Retry and Loss	1/3 of the holdoff users and 1/3 of blocked users exited 2/3 hold-off and 2/3 blocked users waited and retried.
PSwSR (Figure 12)	Yes	Substitute and Retry	User that did not accept price substituted service and retried. User that accepted price and was blocked, substituted service and retried.
PSwSH (Figure 13)	Yes	Substitute and Handoff	Admitted user handed off to the alternate service class with the possibility of a lower price.

The first four experiments (CSwR, CSwRL, CSwR, and CSwSH) did not contain a dynamic pricing function. The first three of these experiments varied in terms of the user response to being blocked by the CAC function in the network. Specifically, all blocked users retried after a wait period in CSwR, shown in Figure 5; some blocked users exited while the remainder retried after a wait period in CSwRL, shown in Figure 6; and all blocked users substituted another service of equivalent QoS in CSwSR, shown in Figure 7. In one additional experiment suggested by this work, admitted users substituted and handed off to an alternate service with the possibility of a lower price as shown in Figure 8. The first two of these were based on the same experiments performed in Hou et al. (2002) and Hou (2003). The third and fourth experiments that introduced substitution were new for this work.

It was expected that the four experiments without dynamic pricing functions (CSwR, CSwRL, CSwSR, and CSwSH) would encounter blocking rates that would increase as the arrival rate increases and would rise significantly higher than 1% as observed in similar experiments in prior work (Hou, 2003; Hou et al., 2002; Manaffar, Bakhshi, & Pilevari, 2008; Olivré, 2004).

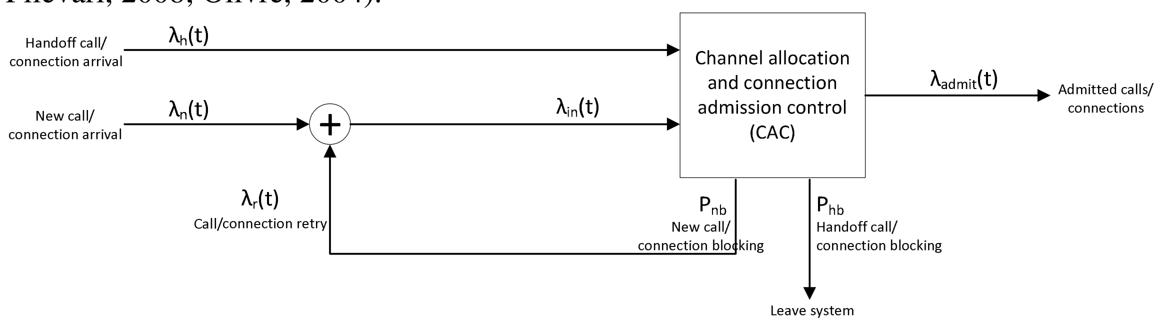

Figure 5. CAC System with Retry (CSwR)

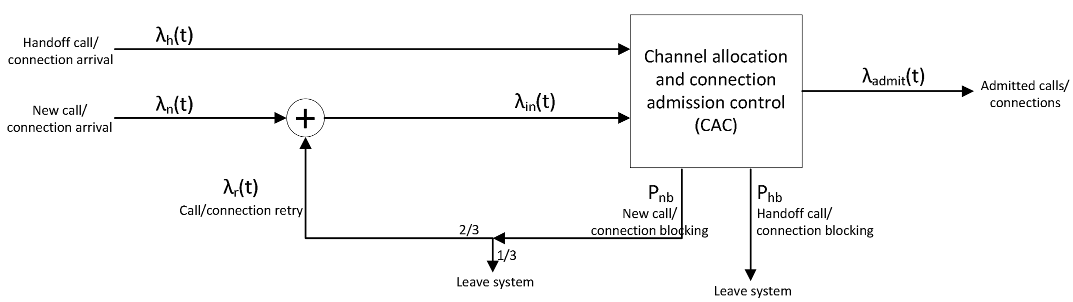

Figure 6. CAC System with Retry and Loss (CSwRL)

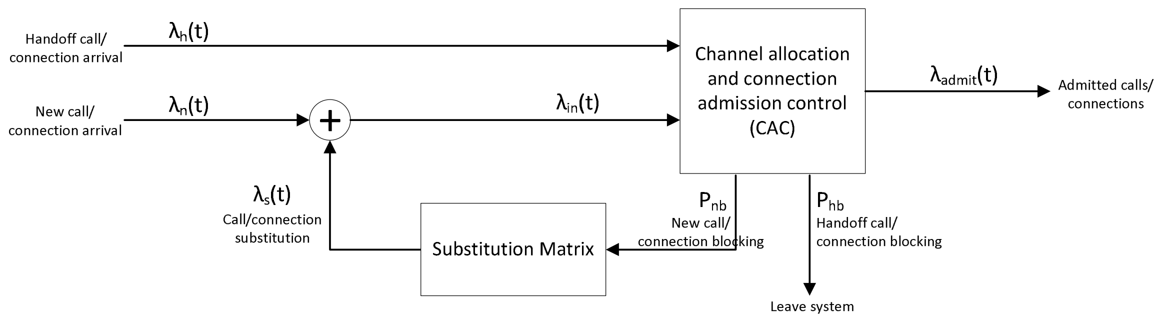

Figure 7. CAC System with Substitute and Retry (CSwSR)

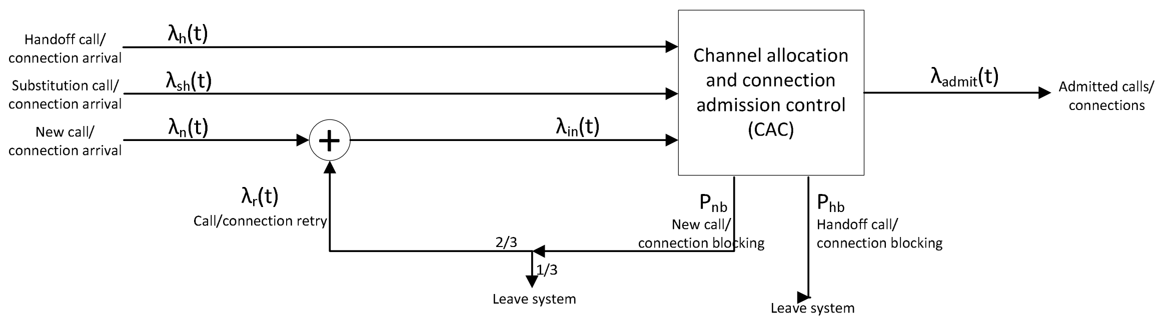

Figure 8. CAC System with Substitute and Handoff (CSwSH)

Five experiments were defined utilizing the dynamic pricing function as part of the CAC system. Four of these experiments differed in terms of the user response to the pricing function: user did not accept price, held off and retried while blocked users exited as shown in Figure 9; hold-off and blocked users retried as shown in Figure 10; some hold-off and blocked users retried while the rest exited as shown in Figure 11; and users that did not accept price, substituted a service with equivalent QoS and retried as shown in Figure 12. The first three (PSwHR, PSwR, PSwRL) were consistent with experiments performed by Hou et al. (2002) and Hou (2003) and the fourth (PSwSR) was new for this work. Finally, a fifth experiment (PSwSH) was suggested by this work based on the ability for a user, during an active session, to be able to substitute service offered by the alternate radio technology and handoff to that service with the possibility of a lower price. This experiment is shown in Figure 13. It was also expected that all of the experiments with dynamic pricing functions (PSwHR, PSwR, PSwRL, PSwSR, and

PSwSH) would also encounter blocking rates that increase as the arrival rate increases but would level off less than or equal to 1% as observed in similar experiments in prior work (Hou, 2003; Hou et al., 2002; Manaffar, Bakhshi, & Pilevari, 2008; Olivré, 2004).

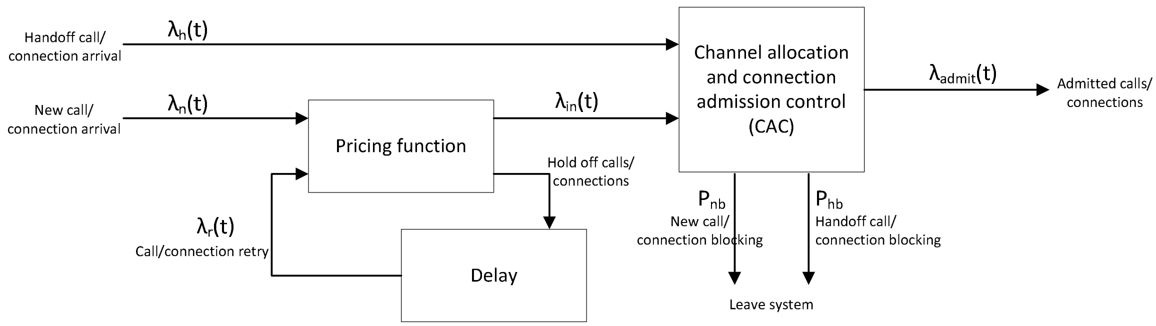

Figure 9. Pricing System with Holdoff and Retry (PSwHR)

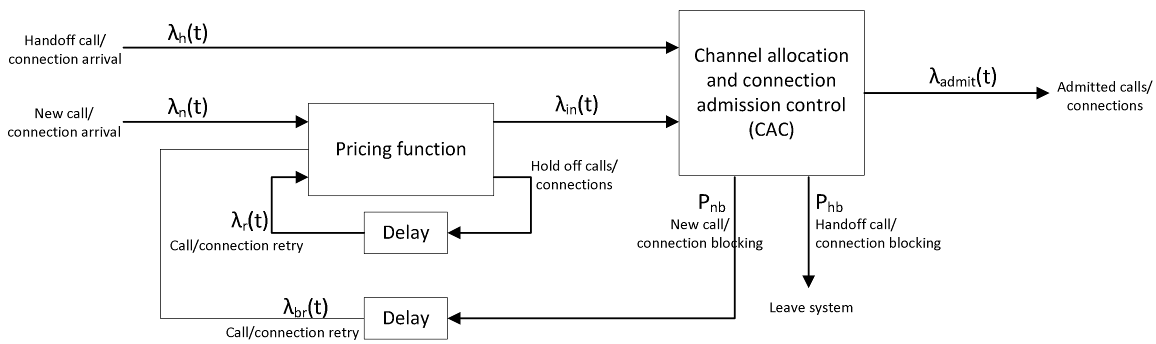

Figure 10. Pricing System with Retry (PSwR)

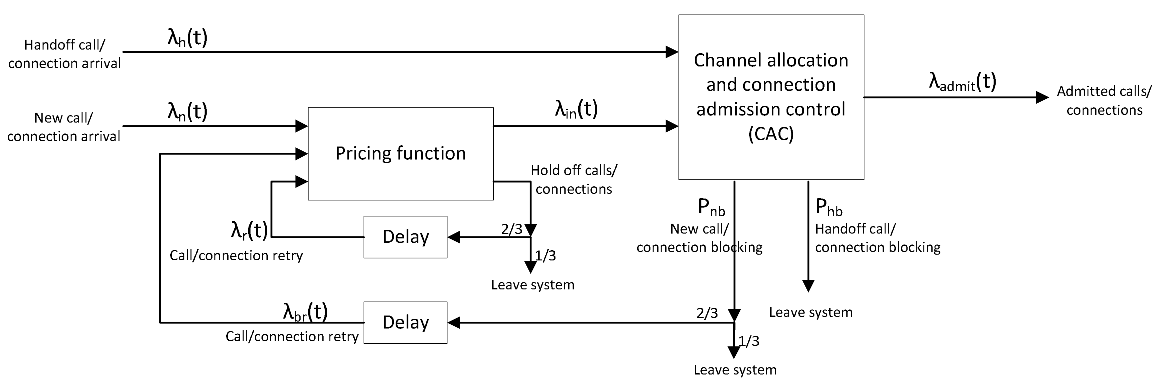

Figure 11. Pricing System with Retry and Loss (PSwRL)

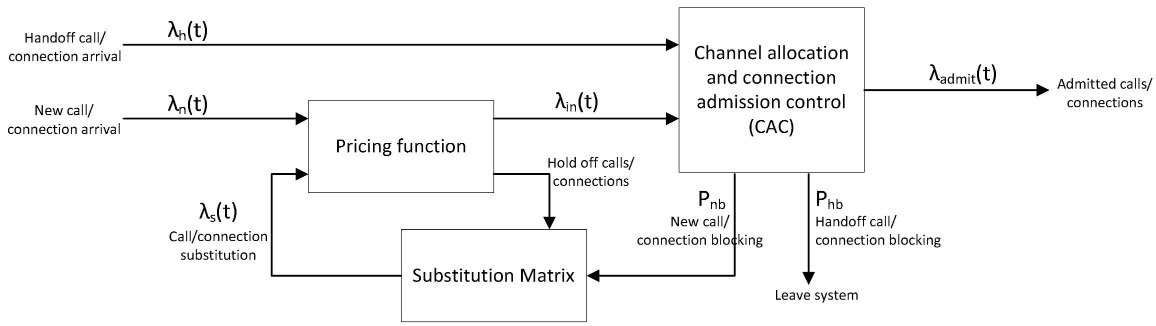

Figure 12. Pricing System with Substitute and Retry (PSwSR)

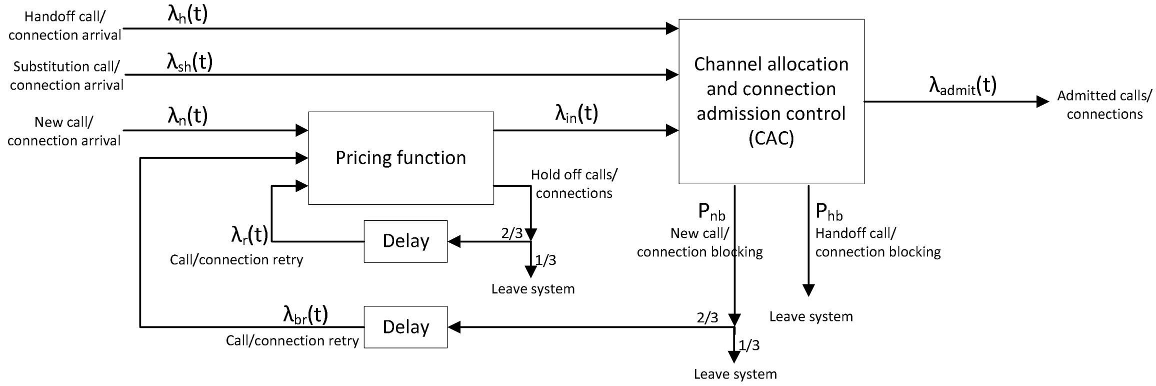

Figure 13. Pricing System with Substitute and Handoff (PSwSH)

Simulation

A discrete-event simulation was developed for the experiments in Table 5. The system simulated a single collocated cell with a representative set of service classes. A random number generator was used to generate a Poisson arrival stream for the new call arrivals. The input arrival rate was varied over a simulated 24 hour period starting from 0 to a point higher than the optimum rate, as shown in Figure 14, thus representing the congestion period as in Hou et al. (2002). This variation represented the demand during a given day with the middle of the day having the highest demand.

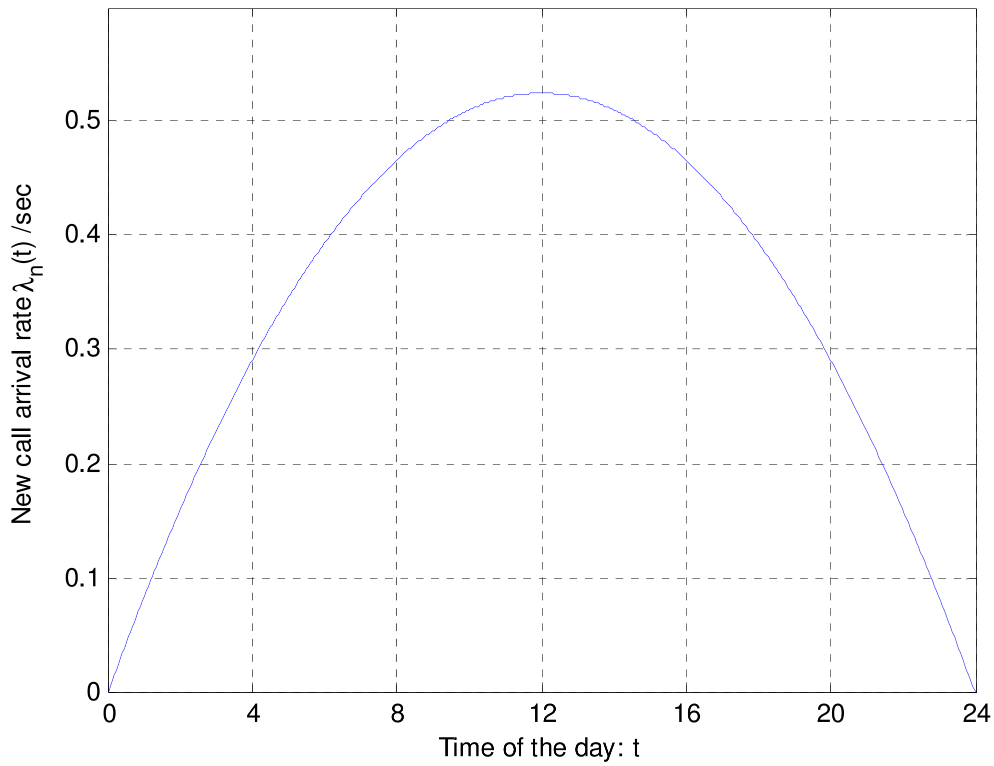

Figure 14. Input new call arrival rate as a function of time.

The simulation utilized the following parameters:

1. As modelled in Al-Manthari et al. (2009), three service classes were modelled, for audio 64 kbps, video 384 kbps, and FTP 128 kbps with a traffic mix of ¼, ¼, and ½, respectively. With a base channel size of 64 kbps, then audio service class required 1 channel, video service class required 6 channels, and FTP required 2 channels. For this work, all three service classes were duplicated in LTE and WiFi with an equal split of input traffic.
2. Each cell was assigned 84 channels with guard channel allocation of 12, each for LTE and for WiFi. All classes within each radio access technology shared the total bandwidth (i.e., 84 channels) and the allocated guard bandwidth was available to be used for handover requests from all service classes serving that

radio access technology (Chao & Chen, 1997). The guard channel allocation (i.e., 12 channels) was chosen as a multiple of the bandwidth requirement for the service class requiring the highest amount (i.e., 6 channels) (Raad, Dutkiewicz, & Chicharo, 2000).

3. Call duration times for new and handoff calls were exponentially distributed with mean ($\frac{1}{\mu}$) 240 seconds.
4. Cell dwell times were exponentially distributed with mean ($\frac{1}{\eta}$) 120 seconds.
5. Arrival of new calls originating in each cell formed a Poisson process with rate $\lambda_n(t)$ as depicted in Figure 14.
6. The new and handoff call probably weights as set to $\alpha = \frac{1}{3}$ and $\beta = \frac{2}{3}$, respectively.
7. The individual user utility for a user of class *k* service, U_k, was the "inelastic" QoS requirement, U_1, as shown in Figure 2.
8. A subset (10%) of admitted users substituted the alternate service class with a service time prior to substitution that was exponentially distributed with mean ($\frac{1}{\mu_s}$) 120 seconds.

The simulation was implemented using Matlab 7 (The Mathworks, 2004) with the add-ons, SimEvents and Simulink (The Mathworks, 2009) on a Windows 7 personal computer. SimEvents provided libraries that allowed for the simulation of discrete-event systems and output statistics that allowed for the measurement of key aspects of the simulation. Coupled with the capabilities offered by Matlab, the call blocking probabilities for the various experiments were determined and analyzed. The new and handoff call blocking rates were plotted through the simulated 24 hour period with the expectation that the call blocking probability would not exceed 1% when dynamic pricing was included.

Resources

As indicated above, the primary experimentation was performed by simulation. The modeling was performed using the student version of Matlab 7 (The Mathworks, 2004) and the simulation was performed using the add-ons, SimEvents and Simulink (The Mathworks, 2009) on a Windows 7 Enterprise 64-bit personal computer.

Following implementation and execution of the simulations, the Matlab 7 environment was used to analyze the data.

Summary

A general model of the dynamic pricing with CAC was presented with a set of user behaviors in response to pricing and blocking conditions. A model with multiple service classes utilizing dual radio technology was defined and a corresponding simulation created with a series of experiments performed incorporating the set of experimental factors: with/without dynamic pricing combined with the set of user response behaviors.

Chapter 4

Results

The overall results demonstrate that the introduction of a dynamic pricing function into a combined LTE and WiFi is successful in preventing congestion by reducing the weighted blocking probability to less than 1%. The system is even successful when the user has the ability to substitute an equivalent service class using the alternate radio access technology. In this scenario, however, there is a delay in the achieving the 1% result as the algorithm takes longer to accommodate the substitutions. The results for the different experiment factors of dynamic pricing function and user behaviors is discussed followed by a discussion of the total utility and total revenue for the different experiments.

Weighted Call Blocking Probability Performance Analysis

When no dynamic pricing algorithm is incorporated in the CAC to control input traffic load, the weighted blocking probabilities can exceed 12%. This is depicted in Figure 15 for Experiment CSwR in which blocked users retry after a wait time and in Figure 16 for Experiment CSwRL in which 1/3 of blocked users leave the system and 2/3 retry after a wait time. Figure 17 and Figure 18 show the traffic loads at different point in the simulation for the CSwR and CSwRL experiments, respectively. Each figure shows the input traffic load $\lambda_n(t)$ over time along with the optimal load constant λ_n^*, the load arriving at the CAC function, $\lambda_{in}(t)$, and the admitted load, $\lambda_{admit}(t)$. One can observe in both scenarios that as the input traffic load exceeded the optimal load in the busy time in the middle of the day, the call blocking increases exceeding the required 1% level. These results, applied to the multi-service calls, heterogeneous radio access technology cellular network, are similar to the results depicted by Hou (2003) for a single service class, single radio access technology cellular network.

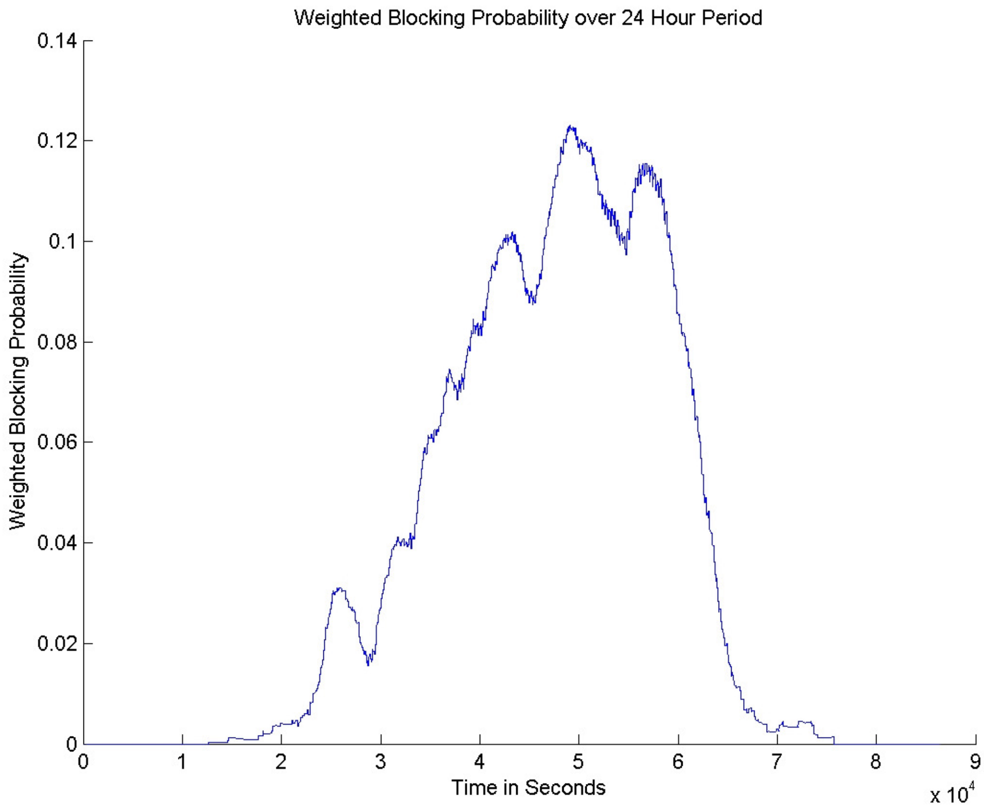

Figure 15. Experiment CSwR: Weighted call blocking probabilities

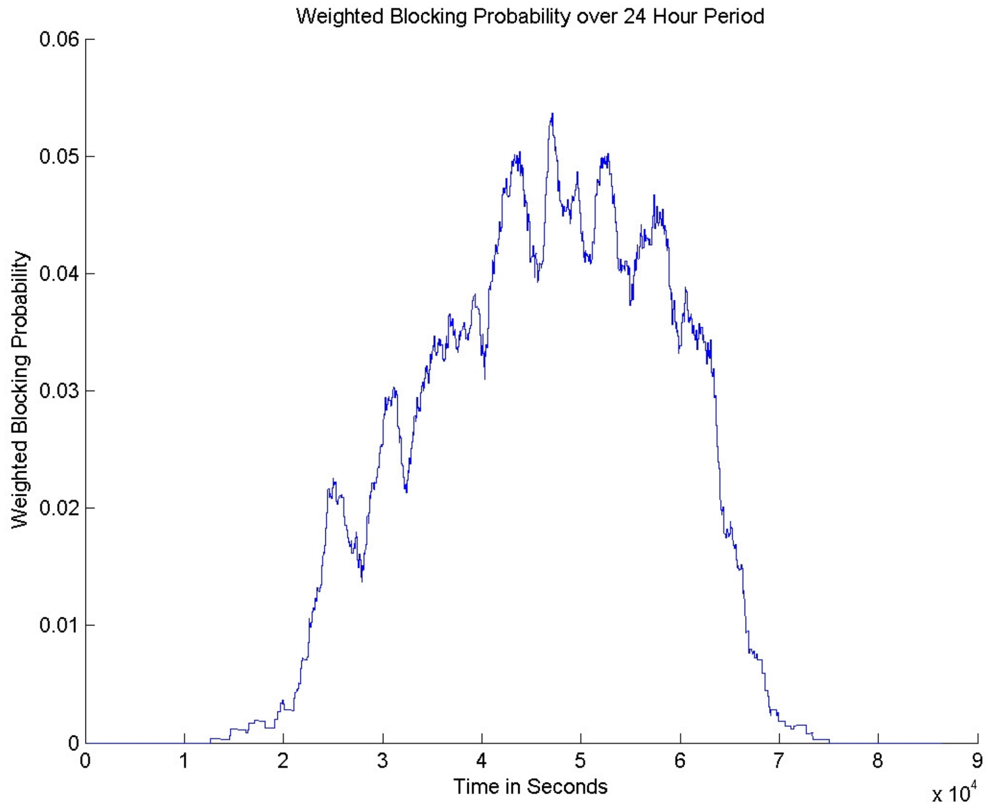

Figure 16. Experiment: CSwRL: Weighted call blocking probabilities

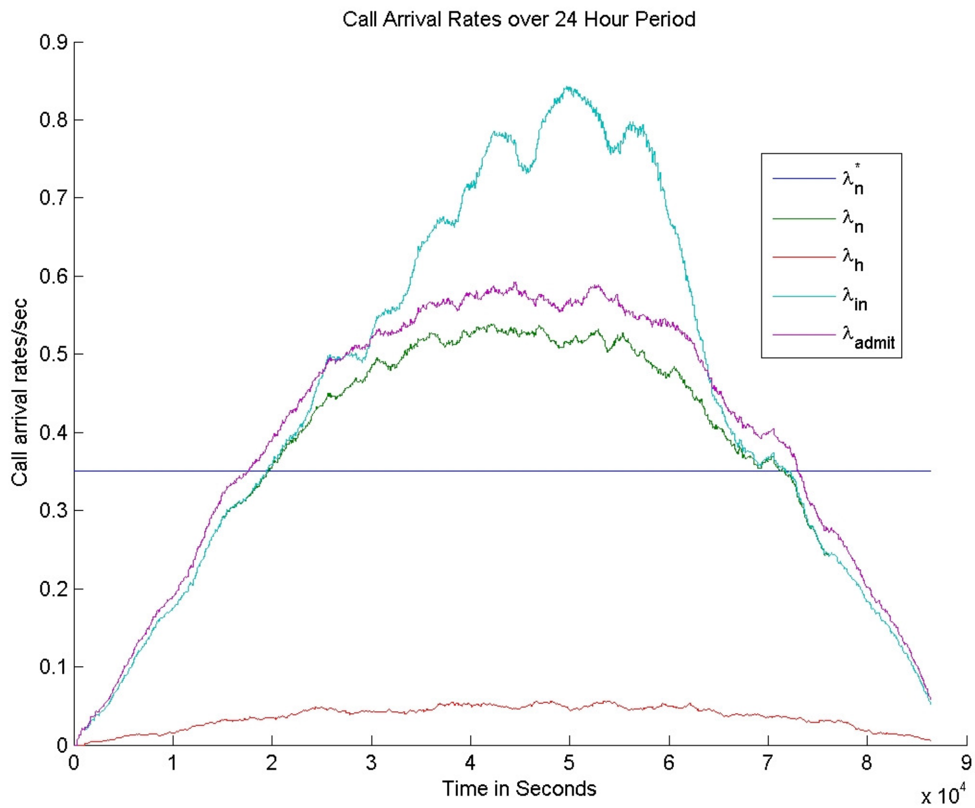

Figure 17. Experiment CSwR: Traffic rates at different points

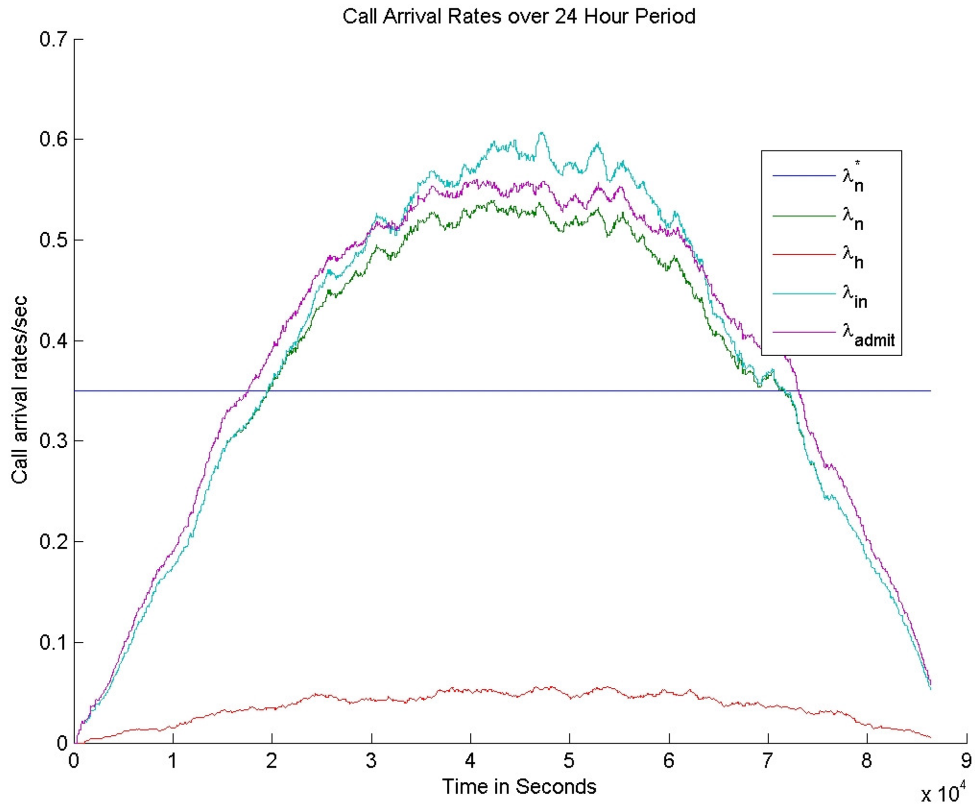

Figure 18. Experiment CSwRL: Traffic rates at different points

When users have the option of substituting a service class using the alternate radio access technology, the results are comparable as depicted in Figure 19, for Experiment CSwSR in which blocked users substitute an equivalent service class and retry and in Figure 20, for Experiment CSwSH in which an admitted user hands off to the alternate service class with the possibility of a lower price. For these scenarios, it can also be seen that as the input traffic load exceeds the optimal load in the busy time, the call blocking increases exceeding the required 1% level. These two experiments show that even though the provisioned bandwidth in the area is increased, simply the introduction of an alternate radio access technology is not sufficient for the operator to meet QoS requirements of their users.

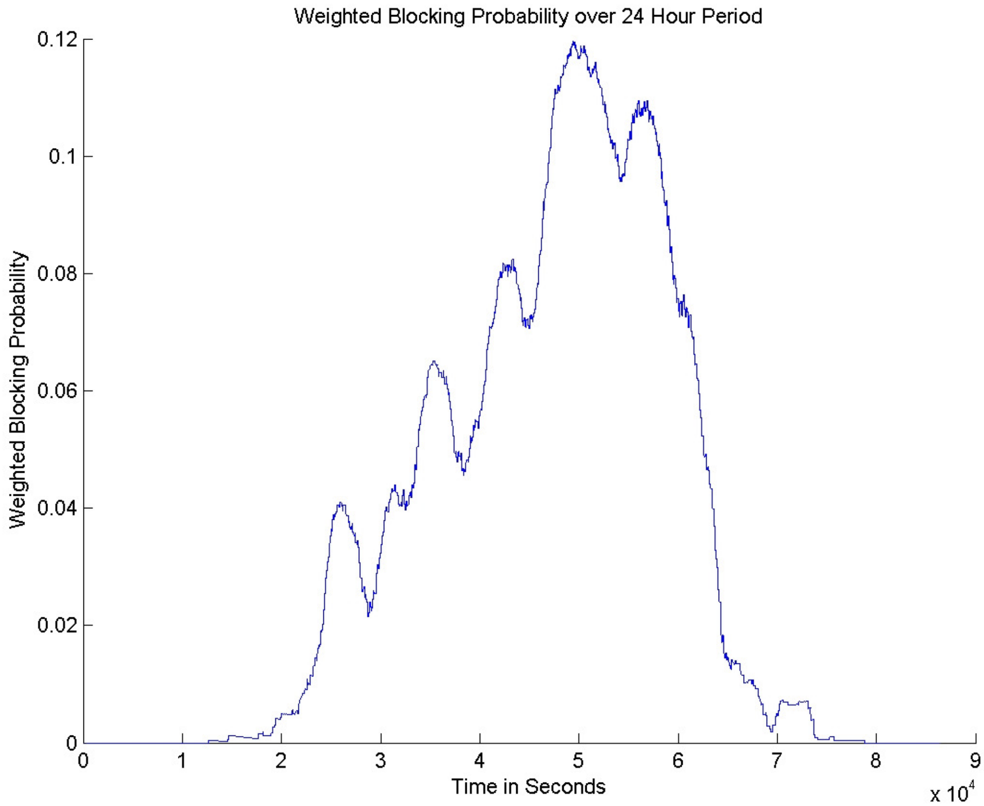

Figure 19. Experiment CSwSR: Weighted call blocking probabilities

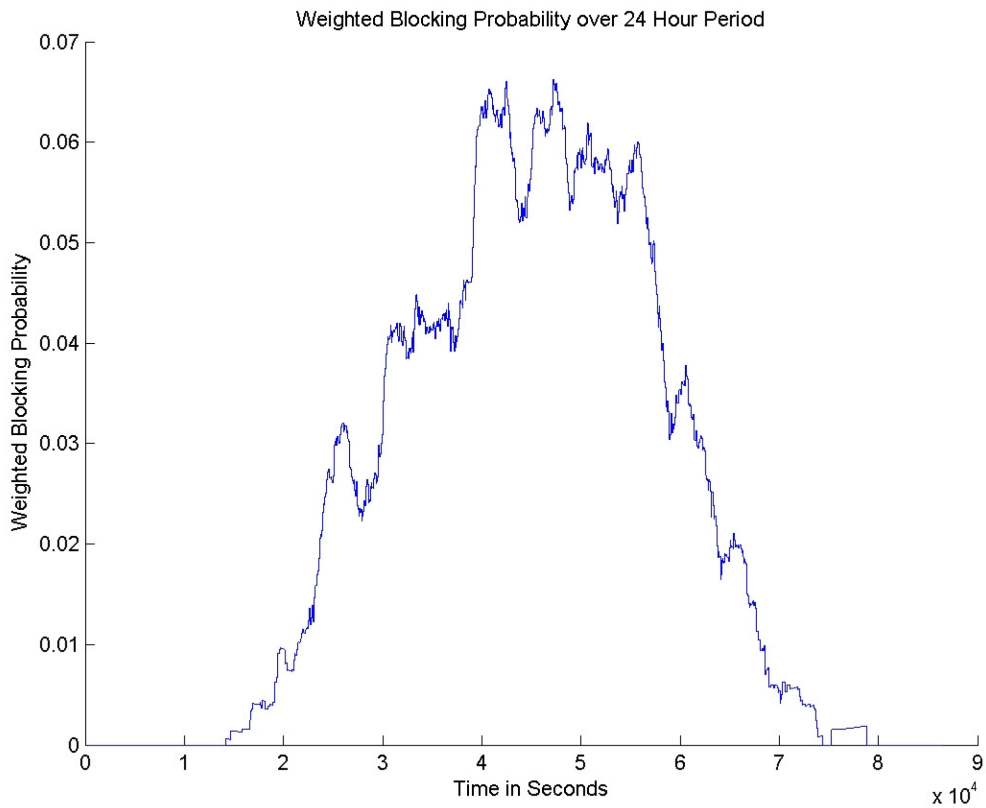

Figure 20. Experiment CSwSH: Weighted call blocking probabilities

The introduction of the dynamic pricing function into the CAC is simulated in the remaining experiments. The price ratio of current price relative to initial price, $p(t)/p_0$, for each of Experiments PSwHR, PSwR, PSwRL is depicted in Figure 21, Figure 22 and Figure 23, respectively. In each, one can observe that when the input traffic load exceeds the optimum input load, the price ratio becomes greater than 1 which demonstrates a price increase at that time relative to the base price. These three experiments all introduce a dynamic pricing function with the only difference being the user behavior in response to the price: user that does not accept price, waits for a while and retry (PSwHR); Holdoff and blocked users will retry after a wait time (PSwR); and only 2/3 of holdoff users and 2/3 of the blocked users retry after a wait time (PSwRL). For each of the cases, while the input traffic load exceeds the optimum load and pricing function activates, one can observe that the traffic load input to the CAC function levels off and approaches the optimum for each of the three experiments as shown in Figure 24, Figure 25 and Figure 26. For each of these three experiments, it is clear that the 1% goal for the blocking probability is met as shown in Figure 27, Figure 28 and Figure 29. These results depict the successful extension of the results achieved by Hou (2003) for a single service class, single radio access technology cellular network to the multi-service, heterogeneous radio access technology cellular network.

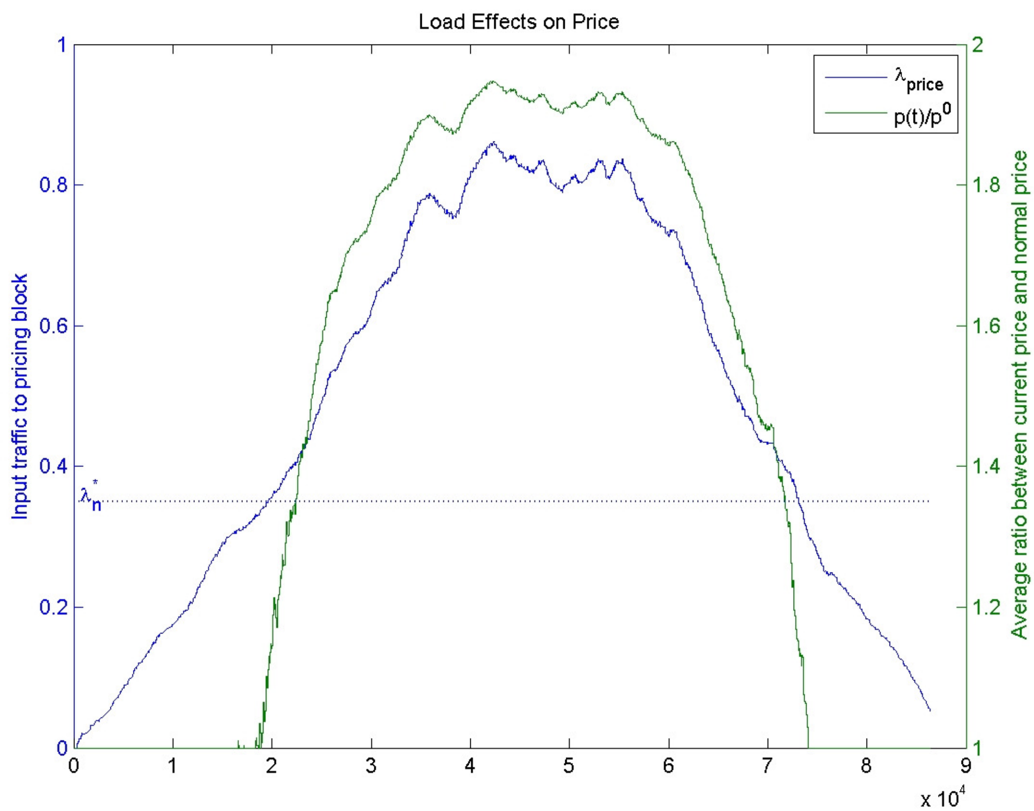

Figure 21. Experiment PSwHR: Load effects on price

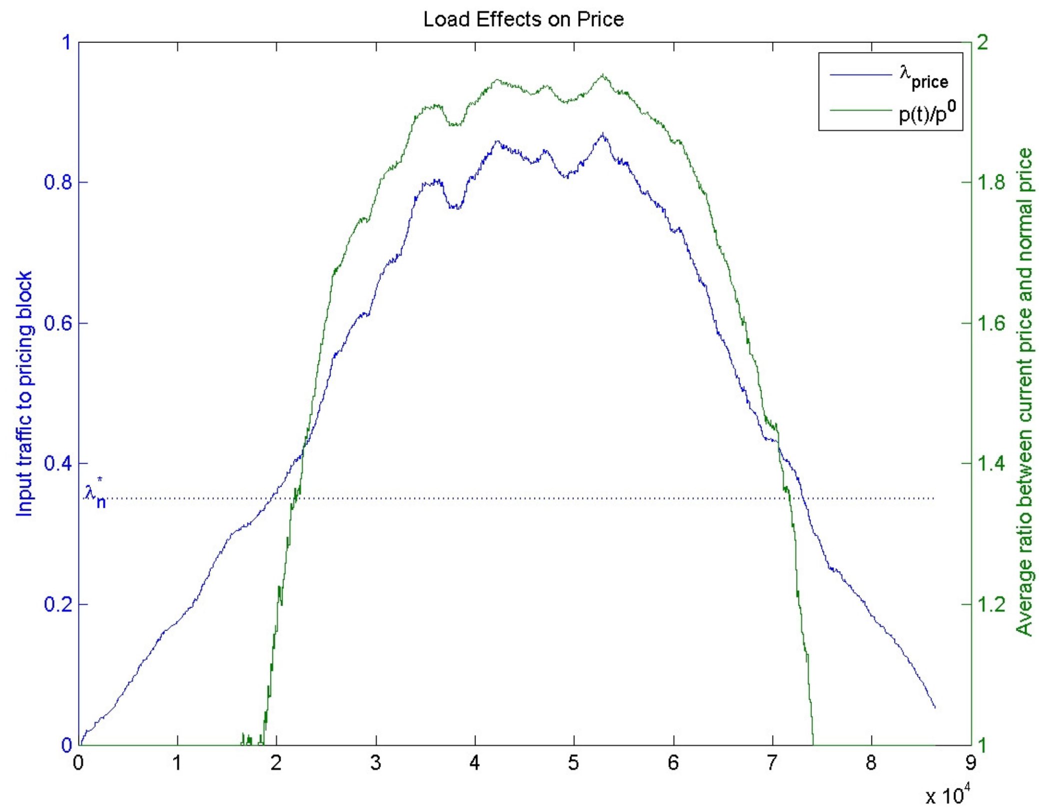

Figure 22. Experiment PSwR: Load effects on price

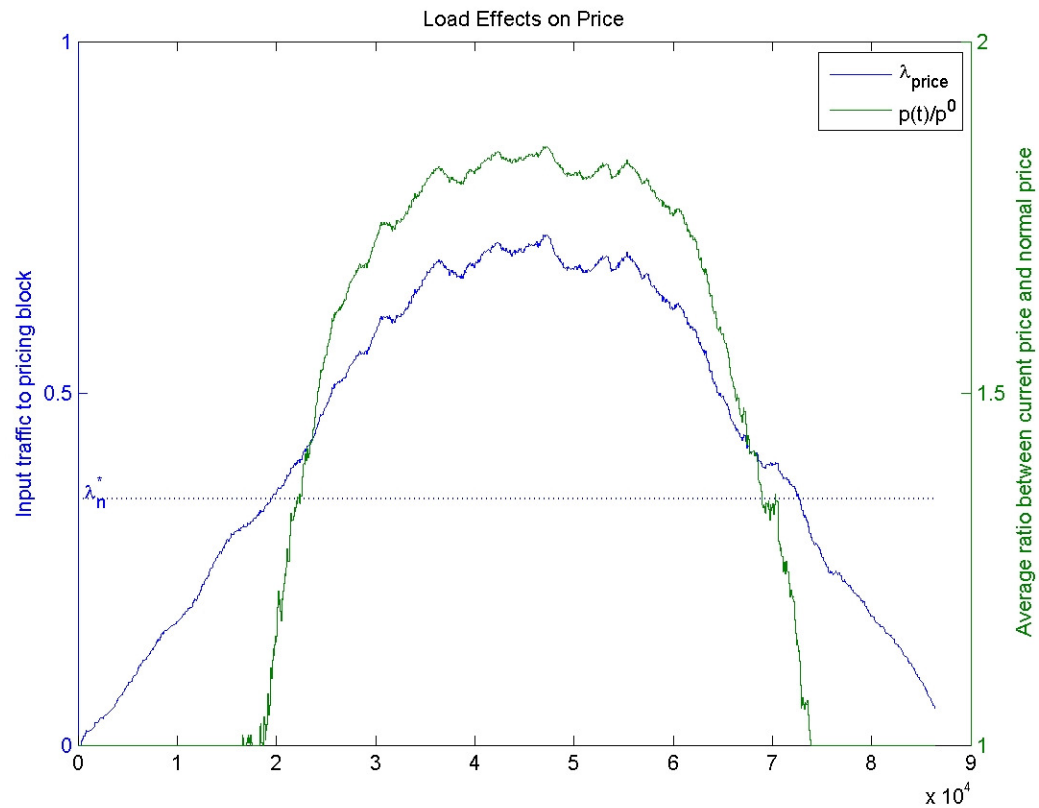

Figure 23. Experiment PSwRL: Load effects on price

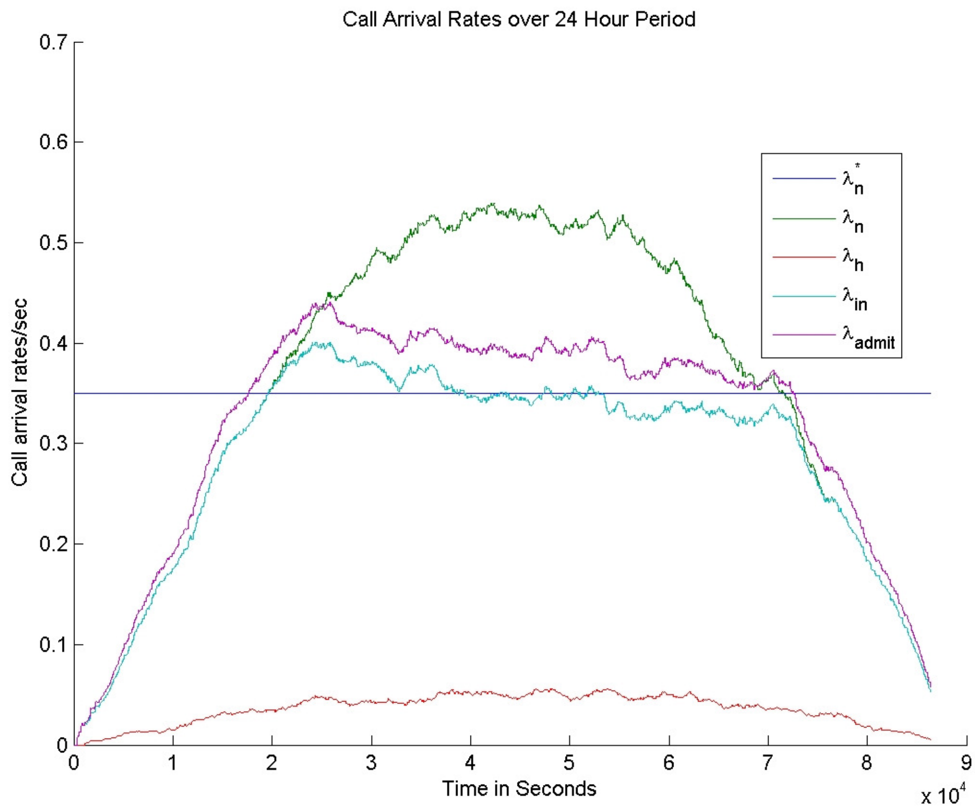

Figure 24. Experiment PSwHR: Traffic rates at different points

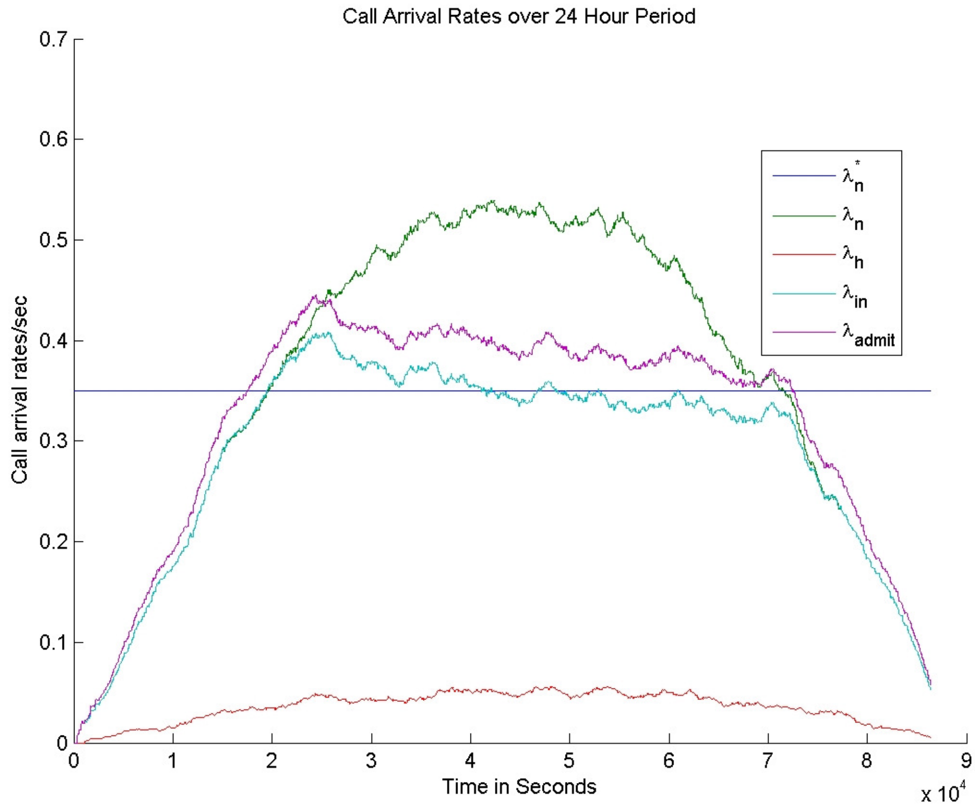

Figure 25. Experiment PSwR: Traffic rates at different points

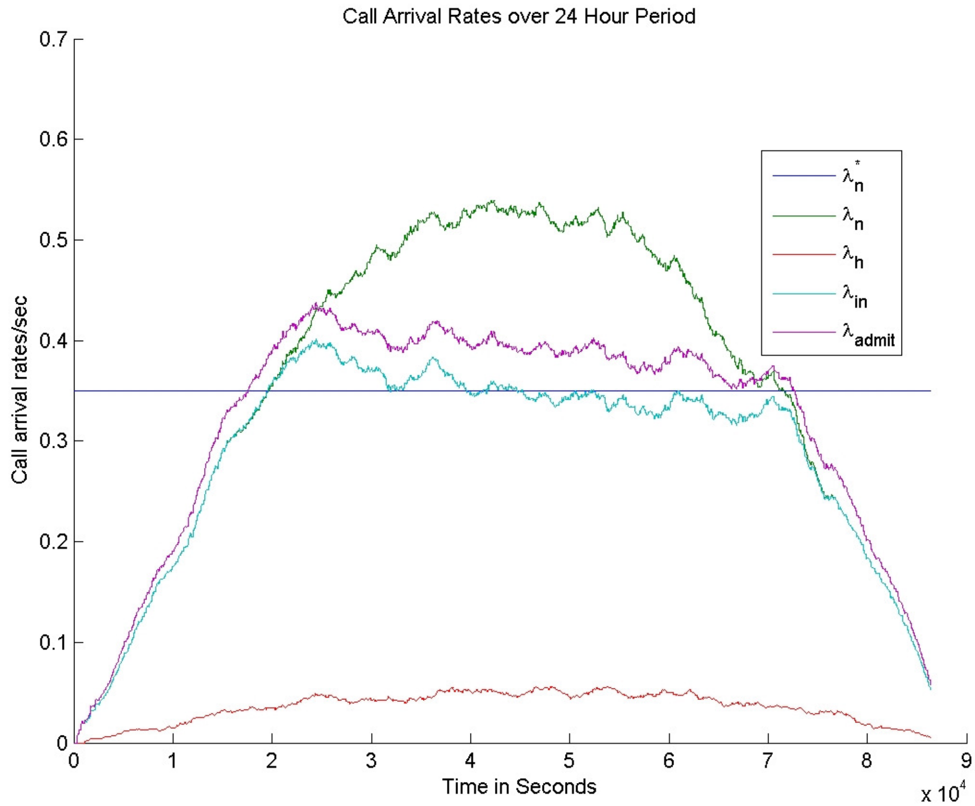

Figure 26. Experiment PSwRL: Traffic rates at different points

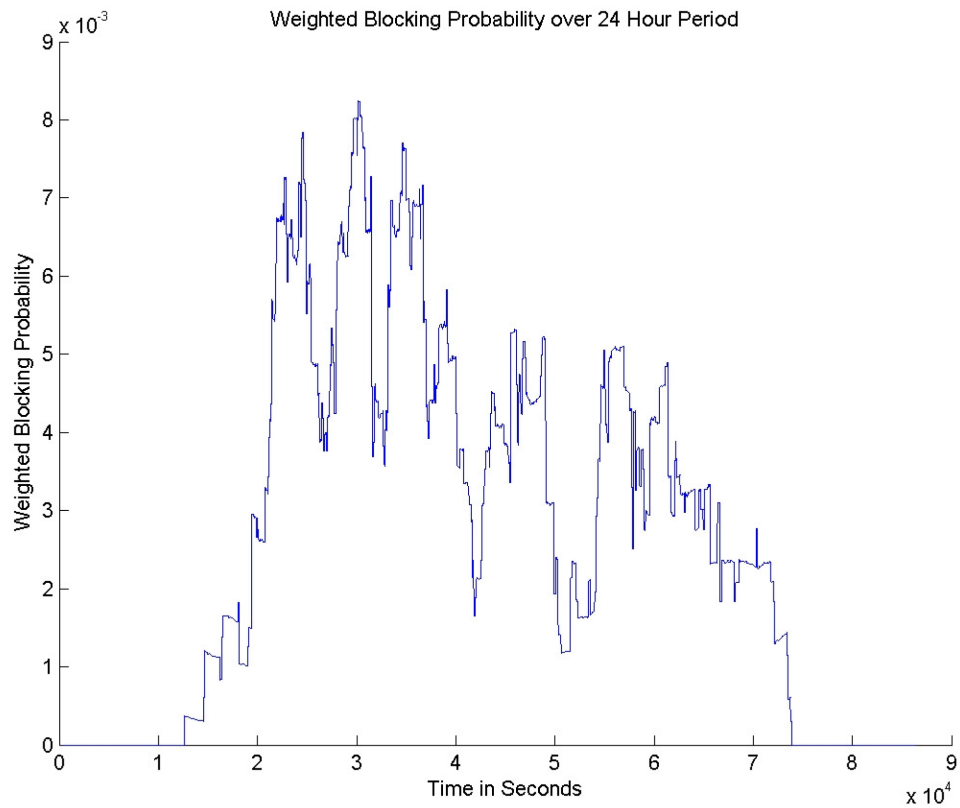

Figure 27. Experiment PSwHR: Weighted call blocking probabilities

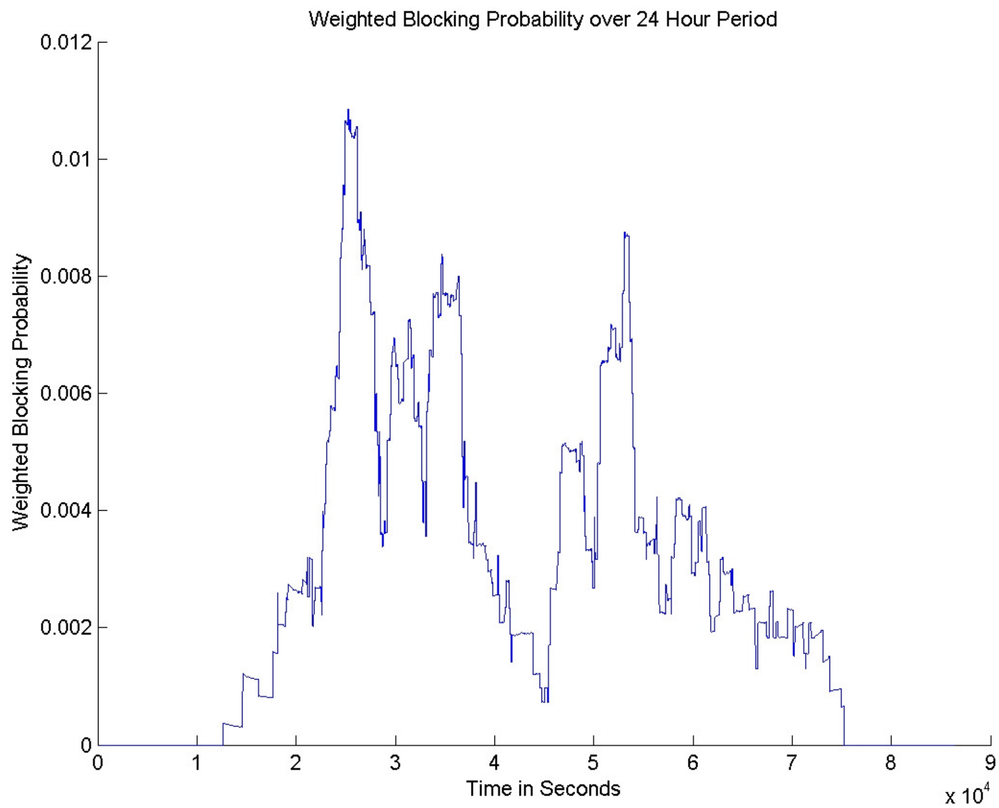

Figure 28. Experiment PSwR: Weighted call blocking probabilities

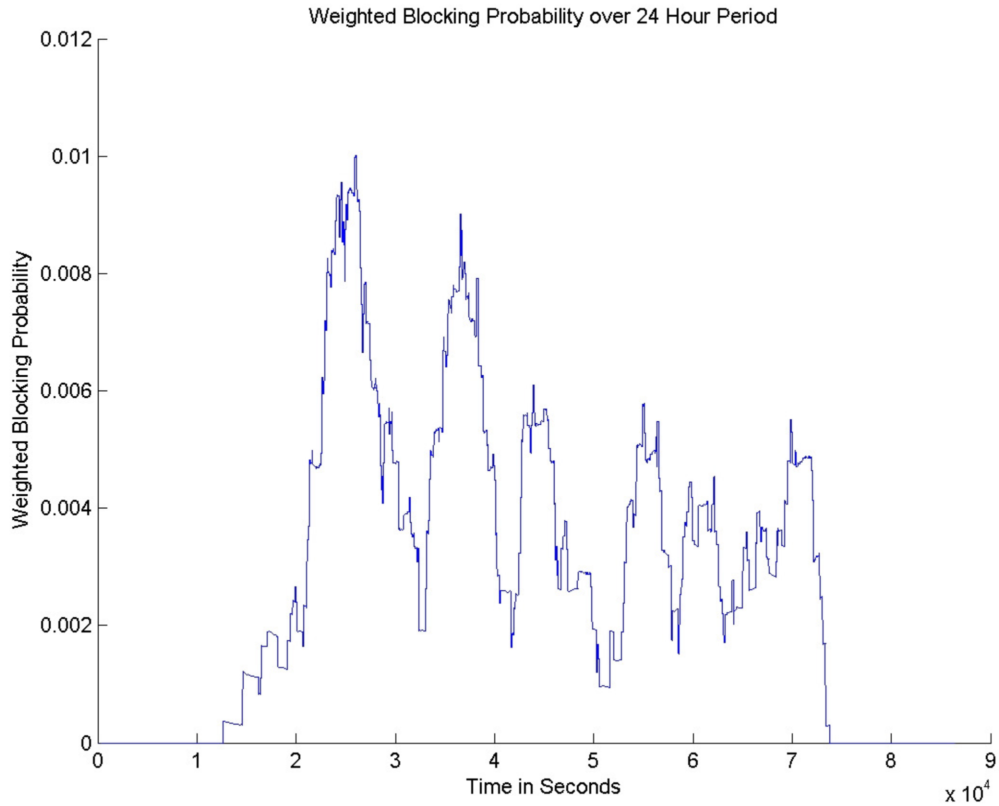

Figure 29. Experiment PSwRL: Weighted call blocking probabilities

Studying the user ability to substitute equivalent service classes introduces an interesting effect into the process. When the user can change service classes from one radio access technology to another, the pricing function based on the load of an individual service class has difficulty achieving the desired goal as quickly. The price ratio, based on traffic load input to the pricing function for Experiment PSwSR, in which holdoff and blocked users both substitute to an equivalent service using the alternate radio access technology, and PSwSH, in which already admitted users substitute an equivalent service class using the alternate radio access technology in the hopes of a lower price, are shown in Figure 30 and Figure 31, respectively. In both cases, the price ratio increases above 1 when the load increases above the optimum load. As in the other three pricing experiments above, the load input to the CAC function also levels out and approaches the optimum as shown in Figure 32 and Figure 33, respectively. However, the user substitutions have an effect on how the call blocking probability drops below the required minimum. It is clear from Figure 34 and Figure 35 that the dynamic pricing function has an immediate effect in causing a significant reduction in call blocking probability, but there are some early spikes above the 1% goal but both do eventually settle below the 1% level.

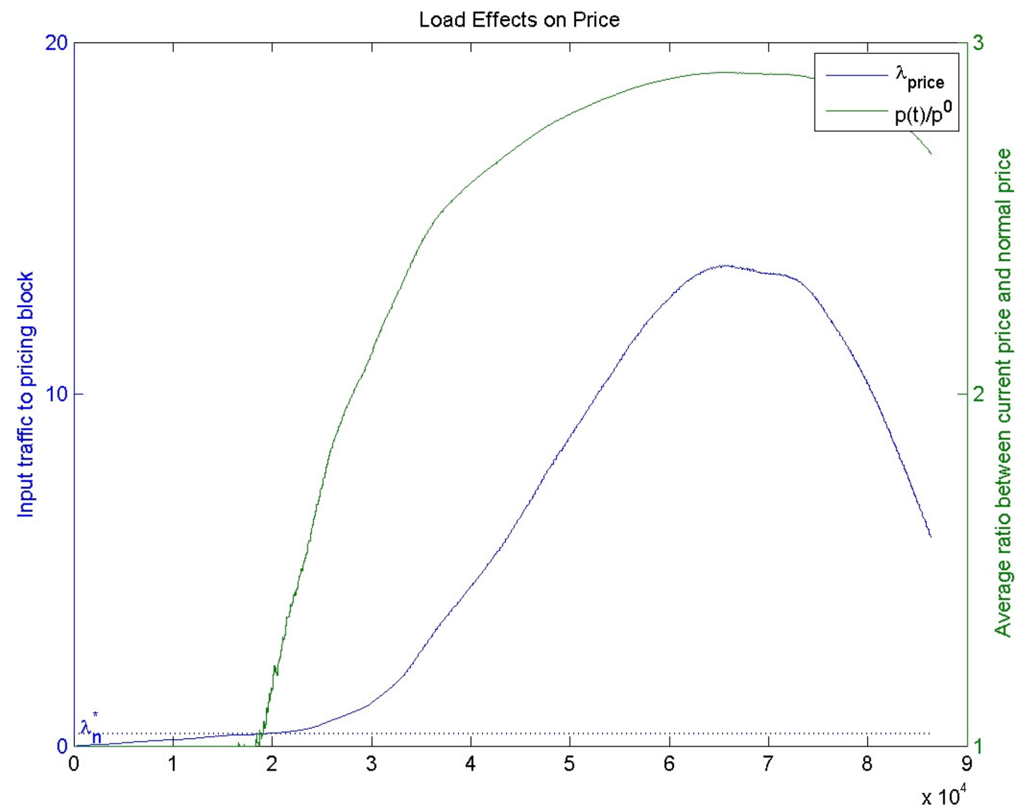

Figure 30. Experiment PSwSR: Load effects on price

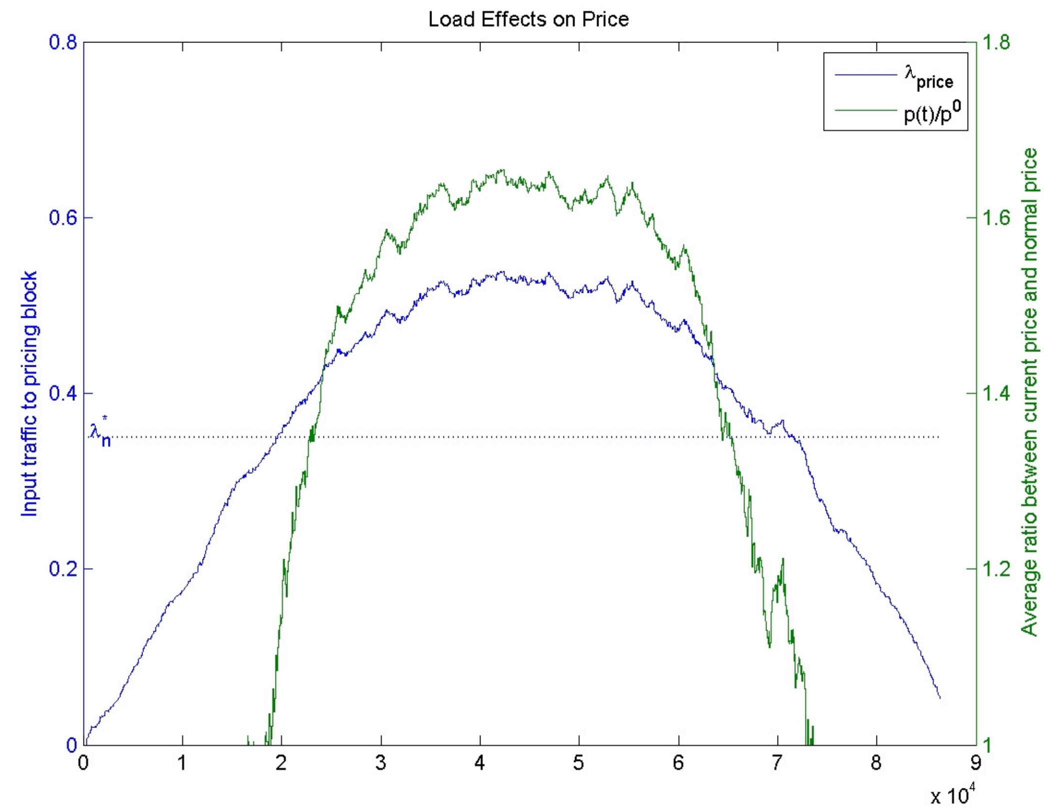

Figure 31. Experiment PSwSH: Load effects on price

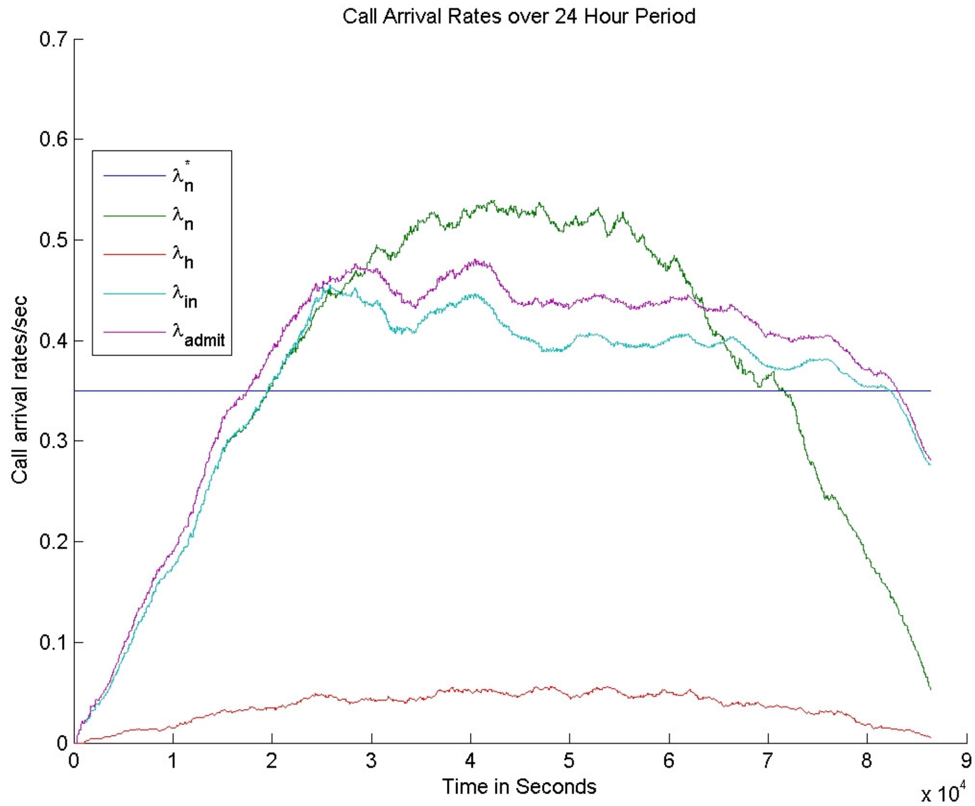

Figure 32. Experiment PSwSR: Traffic rates at different points

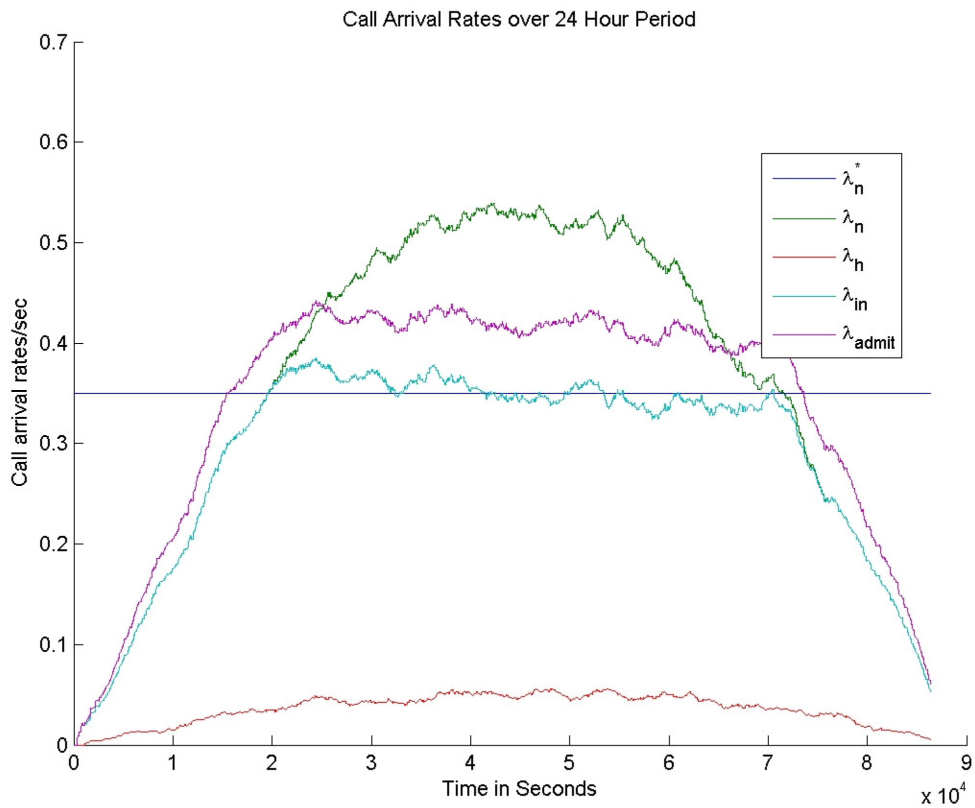

Figure 33. Experiment PSwSH: Traffic rates at different points

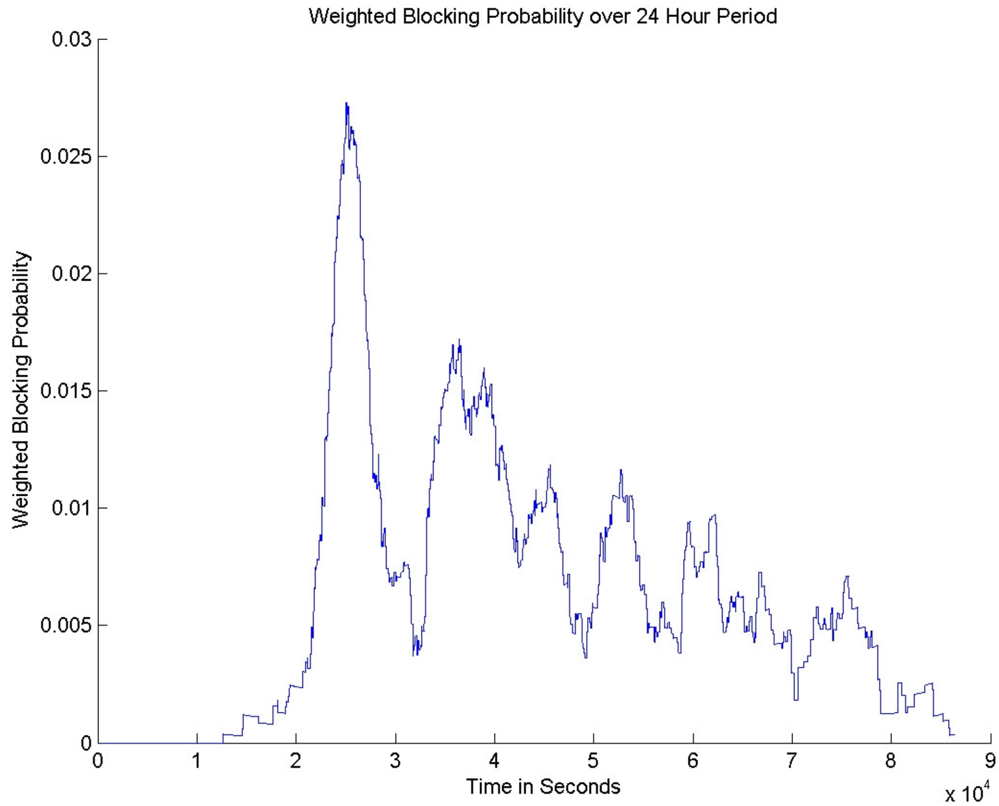

Figure 34. Experiment PSwSR: Weighted call blocking probabilities

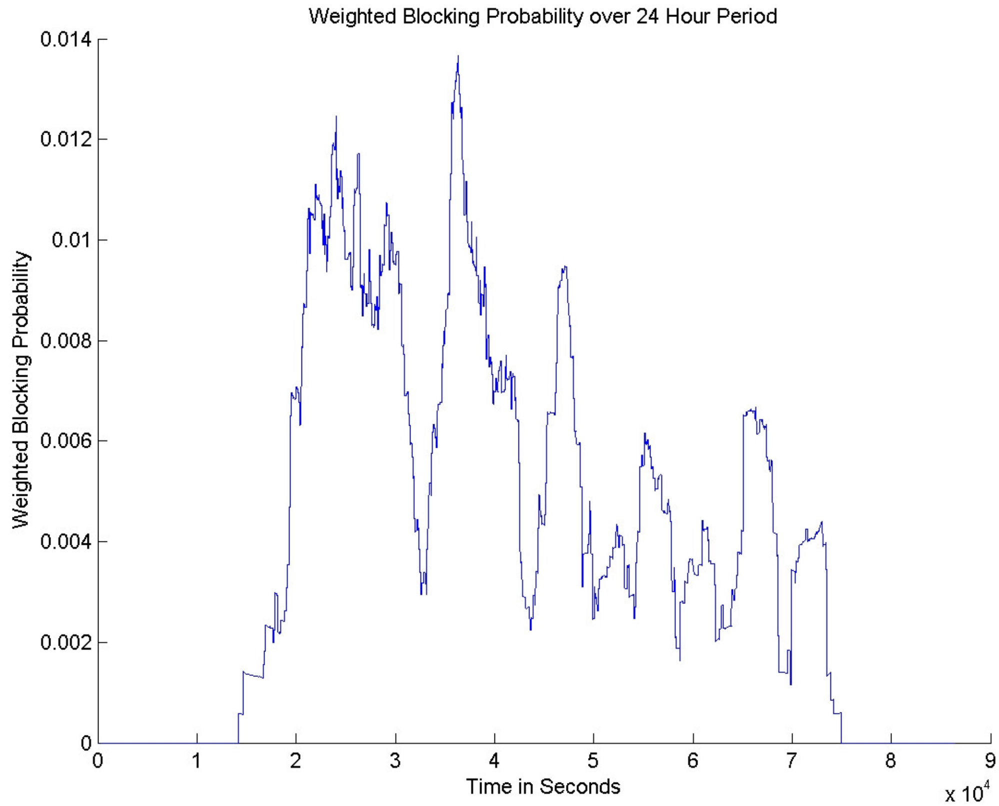

Figure 35. Experiment PSwSH: Weighted call blocking probabilities

Total User Utility and Total Revenue

The total user utility is a measure of how well the network meets the customer needs. The higher the utility, the more effective the network is at meeting the customer requirements. Table 6 lists the total user utility calculated for the single day covered by each experiment. One can observe the introduction of the pricing function doubles the utility of the network, allowing more users into the network.

Table 6. Total User Utility for Each Experiment

	CSwR	**CSwRL**	**CSwSR**	**CSwSH**	**PSwHR**	**PSwR**	**PSwRL**	**PSwSR**	**PSwSH**
Total Utility	11,750	12,430	10,760	10,920	23,550	23,560	23,890	23,550	23,150

The total revenue calculated by users over the 24-hour period is shown in Table 7 for each experiment. The total revenue is increased when the pricing function is integrated with CAC along with the total utility. It is clear that the even though the increase in price deters some users, the increase in total utility indicates that there are still users that are willing to pay the resulting price and unlike CAC alone, the QoS, in the form of blocking probability is assured. The PSwSR experiment yielded a significantly higher revenue than the others. This is due to the fact that the users switching back and forth between the service classes causes some instability in the system and has the effect of increasing the price significantly as shown in Figure 12. The total utility, however, is not significantly increased, so this revenue increase is due more to the instability than to achieving more value of the network. More controlled experiments into variations of service class substitution and its overall effects on the network should be done to determine how a network operator should define the available service classes in order to achieve the most utility and associated revenue from the network.

Table 7. Total Revenue for Each Experiment

	CSwR	**CSwRL**	**CSwSR**	**CSwSH**	**PSwHR**	**PSwR**	**PSwRL**	**PSwSR**	**PSwSH**
Total Revenue ($x10^7$)	1.57	1.51	1.57	1.59	1.89	1.91	1.82	3.37	1.79

Chapter 5

Conclusions, Implications, Recommendations, and Summary

Completed Research

Network operators are struggling to keep pace with the increased demand for network services to meet a wide variety of applications. Some operators are deploying multiple overlapping radio access technologies in an attempt to meet this extra demand. Even with this additional deployment, this over-engineering is not a viable solution and demand will exceed supply leading to congestion and overload conditions (AbuAli et al., 2010; Al-Manthari et al., 2009; Al-Manthari et al., 2011; Courcoubetis & Weber, 2003; Ekstrom, 2009). A solution based on economic principles of dynamic pricing was proposed by Hou et al. (2002) and Hou (2003), who proposed integration of a dynamic pricing function with CAC in order guarantee the QoS, in the form of call blocking probability, in the single service-class cellular network. The work described in this report extended this model in two ways in order to apply it to modern cellular networks: support for multiple service-class networks in order to meet data requirements of different applications and a cooperative deployment of heterogeneous overlapping radio access technologies, like LTE and WiFi. Experiments were performed that enabled the comparison of performance without a dynamic pricing function to performance with a dynamic pricing function. Experiments were also performed demonstrating a variety of behaviors for blocked users and for users that did not accept the current price. In all scenarios, the resulting system that integrated a dynamic pricing function with CAC for multiple service classes across heterogeneous radio access technologies was successful in maintaining the required QoS in the form of a weighted call blocking probability less than 1%.

Equivalent Class Substitution

Of the experiments studied, two of the user behaviors studied are unique to this research work. These two behaviors involved the ability for the user to substitute an equivalent service class served by an alternate radio access technology (e.g., LTE and WiFi) in two scenarios: when the current price was too high and following some period of usage of one class. In the two experiments with the dynamic pricing function integrated with CAC (i.e., PSwSR and PSwSH), the addition of user substitution of an equivalent service class introduced a delay in achieving the 1% blocking probability, but after a short burst over the limit, this, too, was demonstrated successfully. It should be noted that during this time period, when the blocking probability exceeded the 1% target, that users are not experiencing the desired QoS. Further research could be done to explore methods of reducing this over limit time period. The relationship between the traffic loads for these equivalent classes should be studied to determine if this relationship is responsible for introducing the undesirable short term increase in blocking probability. Once this relationship has been validated, then a solution to the problem should also be identified. One option to account for such a relationship could be to incorporate pricing elements based on the load of the equivalent class in the alternate radio access technology in order to balance out the substitution effect.

Lower Class Substitution

As described above, this research work incorporated an equivalent class substitution with favorable, if slightly delayed, results. In some circumstances, as studied by Ozianyi et al. (2008), there are users that will accept a downgraded service in order to maintain their desired price. With the pricing model described in this paper, the focus was only on the load for the current or equivalent classes. When a user that doesn't accept the current price holds off waiting for a better price, that user could alternatively decide to accept a downgraded service class in order to have service immediately at the desired price. This would have the effect of offloading the traffic load for the initial service class, but also introduce an additional traffic load for the downgraded class. This was not included in this study because it was thought to be orthogonal to the alternate radio access technology. In this study, all of the users that did not accept the price, substituted to the alternate class. It is possible that introduction of a substitution matrix that allows for percentages of users to attempt the equivalent class, while another percentage attempt a lower class, the results might improve the delay problem. Another factor that could contribute to this is the fact that the lower class would likely have a lower bandwidth requirement and thus be more likely to be accepted into the system in times of congestion.

Effect of Time on Class Substitution

In the experiments studied, during the onset of congestion, it can be quite some time before the input load returns to a level below the optimum level at which the price is known to satisfy all users. A user that holds off early in the heavy period could be waiting a long time for the price to come back down to an acceptable level. When a user has a significant time to wait, there could be an effect of the duration of the hold time on the user behavior. This can be compared to a negotiation in which time has an effect on the position of one of the negotiating parties that could lead to a loss of bargaining position or a sudden termination (Livne, 1979). For the environment in this paper, the operator's highest goal is to maintain the blocking probability. If the user has a time constraint to accomplish a task requiring the establishment of a data connection, as the hold time increases, the user may choose to accept a higher price in order to secure the service or maintain the price and accept a downgraded service class. On the other hand, if the service is likely to be interactive, then the user may abandon the request when hold off time becomes too great. The time constraint aspects of user behavior may have an impact on the perceived quality of service offered by the operator and could be perceived by the user to be blocked by the network. As a result, the effect of time on the user behavior provides an interesting area of further research.

Summary

Radio technology, in particular, LTE and WiFi, has evolved to support high data rates, but, as described in Chapter 1, given regulatory and technology constraints and the relatively high cost associated with wireless spectrum licensing and utilization, there are still expected to be limitations in the ability to meet the anticipated demands without causing congestion. Hou, Yang, and Papavassiliou (2002) demonstrated that by integrating a dynamic pricing function into the call admission control function of a wireless network serving voice calls, that the resulting network achieved less than 1% blocking probability. The introduction of the dynamic pricing function in which the price increases as the offered input load to the system increases above the derived optimum level, provided an economic solution to the problem of congestion. In Hou et al., users that do not accept the offered price, either leave the system or wait for a lower price, thus reducing the amount of call blocking in the network. Today's data network technology has evolved in two major ways: multiple service classes rather than the single one for voice originally studied by Hou et al. and network overlays of multiple radio access technologies. In order to meet the increasing demands of mobile data users in this environment, this goal documented in this paper was to adapt and apply a dynamic pricing function based on the input traffic load integrated with CAC function to achieve less than 1% blocking probability in a multi-class, dual radio technology network.

In Chapter 2, a review of the literature in the use of dynamic pricing in mobile wireless networks is presented. In this review, the prior art is described starting with the seminal work in this area, Hou et al. (2002), and immediate derivatives of their work all of whose primary limitations is the application to multiple service classes required by data services. Several researchers studied the introduction of dynamic pricing mechanisms into multiple service class data networks with good results that required, however, significant amounts of computation in order to apply. Two groups of researchers (Chen et al., 2010; Ozianyi et al., 2008) introduce connection charges to previously admitted connections which lead to higher network utilization and revenues with the tradeoff of the introduction of a small connection blocking probability, thus not satisfying the 1% target. In the alternate direction, dynamic pricing and heterogeneous wireless networks were also reviewed (Falowo et al., 2010; Wenan et al., 2010), neither of which achieved the new call blocking probability less than 1%.

Chapter 3 provides a generalized model for the dynamic pricing combined with CAC, describes the analytical model that defines the wireless network based on multiple service classes and dual radio access technologies introduced by this work and the experimental design validating this model through simulation. The dynamic pricing function assessed the input traffic load for each service class and when it exceeded the optimum load for the service class, the price was increased according to a standard demand function to control that input load. Nine simulations were performed which varied two factors: dynamic pricing function integration and user behavior. The user behaviors included: blocked users exiting the system or retrying after a delay; blocked users substituting an equivalent service class in the alternate radio access technology and retrying; admitted users handing off to the equivalent service class with possibility of a lower price; user that does not accept price waited and then retried; and users that did not accept price, substituted equivalent service class and retried.

The results of the simulations demonstrating the successful resolution of the blocking problem are presented in Chapter 4. When no dynamic pricing function integrated with CAC, the resulting weighted blocking probability exceeded 12% in some of the experiments. In these scenarios, it was shown that as the input traffic load exceeded the optimal load in the busy time, the call blocking increased, exceeding the required 1%

level. The integration of the dynamic pricing function into the CAC, in all experiments led to the system meeting the 1% goal for the blocking probability. In cases in which class substitution was part of the user behavior, there was an initial increase of blocking probability above the 1% goal, eventually decreasing below the goal.

These anomalies in the case of user substitution, lead to some suggestions for future work in this chapter. Specifically, when equivalent class substitution is supported, should the pricing function be extended to incorporate the load of the equivalent class to account for this related effect. Another area of suggested research is into the user acceptance of a lower service class. Would the delay to acceptable levels be removed when some users select an alternate lower class of service in order to maintain a desired price? One final aspect of class substitution is relevant in terms of the time that the user waits for a better price. In some scenarios, if the delay is too long, the user may be required to accept a higher price for the service, move to the lower class, or exit the system. The combination of these user actions may also have an impact on this delay to acceptable blocking levels.

In conclusion, the research documented in this report presented a dynamic pricing function integrated with CAC for a multiple service class, dual radio access technology network and through simulation demonstrated a blocking probability less than 1%, thus preventing congestion in the wireless mobile network.

References

3GPP. (2009). *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Mobility between 3GPP-Wireless Local Area Network (WLAN) interworking and 3GPP systems (Release 8)* (TS 23.327 v8.4.0).

3GPP. (2010a). *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access (Release 8)* (TS 23.401 v8.9.0).

3GPP. (2010b). *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; IP flow mobility and seamless Wireless Local Area Network (WLAN) offload; Stage 2 (Release 10)* (TS 23.261 v10.1.0).

3GPP. (2011). *SID on user plane congestion management*. 3GPP SA WG2 TD S2-112410. Retrieved from http://www.3gpp.org/ftp/tsg_sa/WG2_Arch/TSGS2_85_XiAN/Docs/S2-112410.zip.

AbuAli, N., Hayajneh, M., & Hassanein, H. (2010). Congestion-based pricing resource management in broadband wireless networks. *IEEE Transactions on Wireless Communications*, 9(8), 2600-2610.

Ahmed, M. H. (2005). Call admission control in wireless networks: a comprehensive survey. *IEEE Communications Surveys & Tutorials, 7*(1), 49-68.

Al-Manthari, B. (2009). *A framework for efficient bandwidth management in Broadband Wireless Access Systems* (Doctoral dissertation, Queen's University (Canada), Canada). Retrieved from ProQuest Dissertations and Theses. (Publication No. AAT NR48208)

Al-Manthari, B., Nasser, N., Abu Ali, N., & Hassanein, H. (2009). Congestion prevention in broadband wireless access systems: An economic approach. *Proceedings of the IEEE Symposium on Computers and Communications (ISCC 2009)*: 606-610.

Al-Manthari, B., Nasser, N., Ali, N. A., & Hassanein, H. (2008, 14-17 Oct. 2008). Efficient bandwidth management in Broadband Wireless Access Systems using CAC-based dynamic pricing. *33rd IEEE Conference on Local Computer Networks (LCN 2008).*

Al-Manthari, B., Nasser, N., & Hassanein, H. (2011). Congestion pricing in wireless cellular networks. *IEEE Communications Surveys & Tutorials*, 13(3), 358-371.

Buddhikot, M. M., Chandranmenon, G., Seungjae, H., Yui-Wah, L., Miller, S., & Salgarelli, L. (2003). Design and implementation of a WLAN/cdma2000 interworking architecture. *IEEE Communications Magazine, 41*(11), 90-100.

Chang, C.-J., Su, T.-T., & Chiang, Y.-Y. (1994). Analysis of a cutoff priority cellular radio system with finite queueing and reneging/dropping. *IEEE/ACM Transactions on Networking, 2*(2), 166-175.

Chao, C.-C., & Chen, W. (1997). Connection admission control for mobile multiple-class personal communications networks. *IEEE Journal on Selected Areas in Communications, 15*(8), 1618-1626.

Chen, F., Ni, R., Qin, X., & Wei, G. (2010). A novel CAC-based network pricing scheme for wireless access networks. *Proceedings of the 2010 IEEE International Conference on Wireless Information Technology and Systems (ICWITS).*

Choi, Y., Ji, H. W., Park, J.-Y., Kim, H.-C., & Silvester, J. A. (2011). A 3W network strategy for mobile data traffic offloading. *IEEE Communications Magazine*, 49(10), 118-123.

Cisco. (2013). Cisco Visual Networking Index: Forecast and Methodology, 2012–2017. San Jose, California: Cisco Systems, Inc.

Courcoubetis, C., & Weber, R. (2003). *Pricing communication networks: economics, technology, and modelling*. Hoboken, NJ: Wiley.

Ekstrom, H. (2009). QoS control in the 3GPP evolved packet system. *IEEE Communications Magazine*, 47(2), 76-83.

Fabrikant, A., Papadimitriou, C., & Talwar, K. (2004). The complexity of pure Nash equilibria. *Proceedings of the thirty-sixth annual ACM symposium on theory of computing*: 604-612.

Falowo, O. E., Zeadally, S., & Chan, H. A. (2010). Dynamic pricing for load-balancing in user-centric joint call admission control of next-generation wireless networks. *International Journal of Communication Systems, 23*(3), 335-368.

Fishburn, P. C., & Odlyzko, A. M. (1998). Dynamic behavior of differential pricing and quality of service options for the Internet. *Proceedings of the first international conference on information and computation economies* (pp. 128-130). Charleston, SC: ACM.

Gizelis, A., & Vergados, D. (2011). A survey of pricing schemes in wireless networks. *IEEE Communications Surveys & Tutorials, 13*(1), 126-145.

Hong, D., & Rappaport, S. S. (1986). Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures. *IEEE Transactions on Vehicular Technology, 35*(3), 77-92.

Hou, J. (2003). *Efficient resource allocation and call admission control in high capacity wireless networks* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Publication No. AAT 3177209)

Hou, J., Yang, J., & Papavassiliou, S. (2002). Integration of pricing with call admission control to meet QoS requirements in cellular networks. *IEEE Transactions on Parallel and Distributed Systems, 13*(9), 898-910.

Hou, J., Yuguang, F., & Akansu, A. N. (2000). Mobility-based channel reservation scheme for wireless mobile networks. *IEEE Wireless Communications and Networking Conference, 2000*.

IEEE (2007). *IEEE standard for local and metropolitan area networks: Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications* (IEEE Std 802.11-2007).

IEEE. (2009). *IEEE standard for local and metropolitan area networks; Part 16: Air interface for broadband wireless access systems* (IEEE Std 802.16-2009).

ITU-R. (2000). *Spectrum Requirements for International Mobile Telecommunications-2000 (IMT-2000)* (Report ITU-R M.2023).

ITU-R. (2006). *Estimated Spectrum Bandwidth Requirements for the Future Development of IMT-2000 and IMT-Advanced* (Report ITU-R M.2078).

Jain, R. (1991). *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling*. New York: Wiley.

Joseph, D. A., Manoj, B. S., & Murthy, C. S. R. (2004). Interoperability of Wi-Fi hotspots and cellular networks. *Proceedings of the 2nd ACM international workshop on wireless mobile applications and services on WLAN hotspots*: 127-136.

Kulavaratharasah, M. D., & Aghvami, A. H. (1999). Teletraffic performance evaluation of microcellular personal communication networks (PCN's) with prioritized handoff procedures. IEEE Transactions on Vehicular Technology, 48(1), 137-152.

Lima, S. R., Carvalho, P., & Freitas, V. (2007). Admission Control in Multiservice IP Networks: Architectural Issues and Trends. *IEEE Communications Magazine*, 45(4), 114-121.

Lin, Y.-B., Mohan, S., & Noerpel, A. (1994). Queueing priority channel assignment strategies for PCS hand-off and initial access. *IEEE Transactions on Vehicular Technology, 43*(3), 704-712.

Livne, Z. A. (1979). *The role of time in negotiations* (Doctoral dissertation). Retrieved from Massachusetts Institute of Technology DSpace@MIT database. (Publication No. 06024309)

Manaffar, M., Bakhshi, H., & Pilevari, M. (2008). A new dynamic pricing scheme with call admission control to reduce network congestion. *Proceedings of the 22nd International Conference on Advanced Information Networking and Applications (AINAW 2008)* (pp. 347-352). Piscataway, NJ: IEEE.

Niyato, D., & Hossain, E. (2005). Call admission control for QoS provisioning in 4G wireless networks: issues and approaches. *IEEE Network*, 19(5), 5-11.

Olivré, A. (2004). *Call admission control and dynamic pricing in a GSM/GPRS cellular network* (Master's thesis, University of Dublin, Trinity College, Dublin, Ireland). Retrieved from https://www.scss.tcd.ie/publications/tech-reports/reports.05/TCD-CS-2005-16.pdf (TCD-CS-2005-16)

Ozianyi, V. G., Ventura, N., & Golovins, E. (2008). A novel pricing approach to support QoS in 3G networks. *Computer Networks, 52*(7), 1433-1450.

Raad, R. S., Dutkiewicz, E., & Chicharo, J. (2000). Connection admission control in micro-cellular multi-service mobile networks. *Proceedings of the Fifth IEEE Symposium on Computers and Communications* (pp. 600-606). Piscataway, NJ: IEEE.

Salkintzis, A. K., Fors, C., & Pazhyannur, R. (2002). WLAN-GPRS integration for next-generation mobile data networks. *IEEE Wireless Communications, 9*(5), 112-124.

Sandvine. (2013). Global Internet Phenomena Report - 2H 2013. Waterloo, Ontario, Canada: Sandvine Incorporated.

Soldatos, J., Vayias, E., & Kormentzas, G. (2005). On the building blocks of quality of service in heterogeneous IP networks. *IEEE Communications Surveys & Tutorials*, 7(1), 69-88.

The Mathworks, Inc. (2004). *Matlab 7: The Language of Technical Computing*. Retrieved from http://www.mathworks.com/academia/student_version/r2011a_products/matlab.pdf

The Mathworks, Inc. (2009). *SimEvents 3.0: Model and simulate discrete-event systems*. Retrieved from http://www.mathworks.com/academia/student_version/r2011a_products/simevents.pdf

Wenan, Z., Xiaoli, A., Bing, X., & Junde, S. (2010). A novel QoS and profit oriented dynamic pricing and admission control mechanism. Paper presented at the 2010 6th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM).