On the Long-term Wireless Network Deployment Strategies

WU QI MING

NATIONAL UNIVERSITY OF SINGAPORE

2007

On the Long-term Wireless Network Deployment Strategies

WU QI MING

(B.Eng, SHANGHAI JIAO TONG UNIVERSITY)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ENGINEERING

DEPARTMENT OF ELECTRIC AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2007

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my supervisor, Dr. Chew Yong Huat, who is from the Institute for Infocomm Research (I^2R). Thanks for his invaluable guidance, support, encouragement, patience, advice and comments throughout my dissertation. His rigorous academic attitude has imbued a deep sense of value in me. Without his help, I might not be able to complete this thesis. It is him who encouraged me to complete my research using my after office hour when I was about to give up.

I also want to thank Dr. Yeo Boon Sain, who was previously also from I^2R for the encouragement and inspiration given to my research topic.

I also want to thank my wife, who is always by my side, supporting and encouraging me to go through all the hard times.

Last but not least, I want to thank my parents. Their love and never ending support are what I treasure the most.

TABLE OF CONTENTS

LIST OF NOTATIONS

g	a factor to reserve some bandwidth in the APs
A	set of candidate sites
a_i	whether AP i should be activated. $a_i = 0$ if the i^{th} AP is not activated, $a_i = 1$ if the AP is activated
B	channel capacity
C_A	the hardware cost
C_c	the running cost of a channel
$C_{I\,i}$	the initialization cost such as installation
c_{ik}	the assignment parameter indicates that channel k of AP i is being activated
$C_{M\,i}$	the maintenance cost of AP i and
D_{ij}	the equivalent distance between AP i and DP j
f	the amount of attenuation (in dB) due to the presence of floors in the propagation path
$f_{i_1 i_2}$	$f_{i_1 i_2} = 0$ is used to indicate no interfere between APs i_1 and i_2, For $f_{i_1 i_2} = 1$, the relationship $r_{i_1} + r_{i_2} \le 2R_{\max}$ gives no effect
i	candidate site index
j	demand point index
k	channel number
l_p	number of feasible solutions for period p

M	set of demand points
m	the number of walls
n	the number of floors
P	the utilization charge for per unit demand traffic
$R_{\max}$	the maximum coverage radius of an AP
r_i	the equivalent coverage radius (i.e., attenuation due to the presence of obstacles has been compensated by an additional distance) of the AP for its transmission power
T_j	the demand traffic at DP j
t_{ij}	traffic from the i^{th} AP to DP j
w	the amount of attenuation (in dB) due to the presence of walls in the propagation path
x_{ij}	the assignment parameter denote the link between DP j and candidate AP site i

LIST OF ABBREVIATIONS

AMPL	Modeling Language for Mathematical Programming
AMPS	Advanced Mobile Phone System
ANSI	American National Standards Institute
AP	Access Point
BPSK	Binary Phase Shift Keying
BS	Base Station
CCK	Complementary code keying
CDMA	Code Division Multiple Access
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
DCF	Distributed Coordination Function
DIFS	DCF Inter frame Space
DP	Demand Point
DSSS	Direct Sequence Spread Spectrum
FDMA	Frequency Division Multiply Access
GLPK	GNU Linear Programming Kit
GNU	GNU's Not Unix
GPL	General Public License
GSM	Global System for Mobile Communications
IEEE	Institute of Electrical and Electronics Engineers

ILP	Integer Linear Programming
LAN	LocalAreaNetwork
MAC	Media Access Control
MAN	MetropolitanArea Network
MILP	MixedInteger Linear Programming
NP-HARD	Nondeterministic Polynomial-time hard
OFDM	Orthogonal Frequency Division Multiplexing
PCF	Point coordination function
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase-shift keying
RTS/CTS	Request-To-Send/Clear-To-Send
SP	Service Provider
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunications System
VPN	Virtual Private Network
WAN	Wide Area Network
WCDMA	Wideband Code Division Multiple Access
WEP	WiredEquivalent Privacy
WiFi	Wireless Fidelity
WLAN	Wireless LocalArea Network

LIST OF FIGURES

LIST OF TABLES

SUMMARY

The deployment of wireless networks needs to consider both the cost and system performance metrics. The design objective is to decide the optimal placement of access points (or base stations) and to assign the available radio resources to respective traffic demand points with guaranteed performance, while keeping the deployment cost at its minimum. The optimization problem can become quite complicated when multiple performance metrics need to be satisfied concurrently. Most of the reported works formulate the problem using mixed integer linear programming (MILP) but with different objective functions. Normally these works assumed that demand traffic do not vary with time. However, we feel that it would be better to adopt a design which can achieve long-term optimal rather than just at the instant of deployment. In this thesis, we set up a platform to look into the deployment of wireless networks which is able to optimize the profit generated over multiple periods each with different spatial traffic demands. Given a set of candidate sites, we first derive the placement and compute the transmission power of the access points to support a given spatial traffic demand over a specific period of time. The problem was also formulated using a mixed integer linear programming model, both for the FDMA based and CSMA/CA based networks. Adjustable transmission range is made possible through power control to minimize the amount of interference among neighboring access points. With the knowledge on the projected demand traffic in

subsequent periods, algorithms to maximize the long term profit are developed, both when the projected traffic are probabilistic and deterministic.

CHAPTER 1

INTRODUCTION

1.1. Evolution of Mobile communications

Wireless technology has been developed for more than a century since Guglielmo Marconi invented the world's first wireless telegraph in 1896. Today wireless communication devices and technologies have penetrated our daily lives and it will surely continue to be so in the next decade.

From 1896, wireless communication technologies have gone through various evolutions. When communication satellites were first launched in the 1960s, those satellites could only handle 240 voice circuits. Today, satellites carry about one-third of the total voice traffic and all the television signals between countries. The cellular or mobile telephones which are the modern equivalent of Marconi's wireless telegraph, can now offer very reliable two-party with two-way communication even under high user mobility. The first–generation of wireless phones used the analog technology. The dominant first-generation digital wireless network in North America was the Advanced Mobile Phone System (AMPS). The network devices were bulky and coverage was patchy, but they successfully demonstrated the inherent convenience to perform communications between two parties [1]. The current or second-generation of wireless devices are using digital technology instead of analog. The existing deployed second-generation wireless systems are the GSM [2], PCS IS-136 and PCS IS-95. Cellular systems such as GSM are optimized for wide-area coverage, and can provide bit rates around 200kbps in each carrier frequency. The third-generation of cellular systems, also known as the Universal Mobile Telecommunications Systems (UMTS [3]), aims to deliver data rates of 384 kbps for high mobility users and up to 2 Mbps for low mobility users. The UMTS standard was based on WCDMA technology. Another UMTS proposal is based on the CDMA2000 by the United States, which is compatible with IS-95 CDMA.

With the booming use of Internet in the recent decades, users are demanding for more bandwidth to transmit multimedia traffic. Service providers therefore have the urge to develop wireless networks which can support higher data transmission rate in order to meet the users' needs. Higher data rate systems are now achievable with the development of broadband wireless technology. The two key factors to make wireless services to a success and become popular are the convenience to access (i.e., good coverage and reliable) and the lower development costs (i.e. cheap). The low development cost can be achieved now with today's wireless technologies since operators can provide the usual telecommunication services with low cost wireless devices. The fact that services are possible for high mobility users has provided us with a convenient way of performing communications. Furthermore, standardization is also necessary to ensure interoperability between devices developed by different vendors. There are many initiatives in developing broadband wireless standards for different applications, from the wireless LAN to the small wireless home network. Their data rates vary from 2 Mbps to well over 100 Mbps. Many of these technologies are available now and more will become available in the next several years. Among these standards, Wi-Fi seems to get much more attention in the recent years. Wi-Fi (IEEE 802.11) denotes a set of wireless LAN/WLAN standards developed by the working group 11 of the IEEE LAN/MAN Standards Committee (IEEE 802). The term 802.11x is also used to denote the set of amendments to the standard.

The original version of the standard IEEE 802.11 [4] released in 1997 specifies two raw data rates of 1 and 2 megabits per second (Mbps) to be transmitted via infrared (IR) signals or by either frequency hopping or direct-sequence spread spectrum in the Industrial Scientific Medical (ISM) frequency band at 2.4 GHz. The 802.11a amendment to the original standard was ratified in 1999. The 802.11a standard uses the same core protocol as the original standard, operates in 5 GHz band, and uses a 52-subcarrier OFDM with a maximum raw data rate of 54 Mbps, which yields realistic net achievable throughput in the mid-20 Mbps. The data rate can be reduced to 48, 36, 24, 18, 12, 9 then 6 Mbps if required. The 802.11b amendment to the original standard was ratified in 1999. 802.11b has a maximum raw data rate of 11 Mbps and uses the same CSMA/CA media access method defined in the original standard. Due to the CSMA/CA protocol overhead, in practice the maximum 802.11b throughput that an application can achieve is about 5.9 Mbps using TCP and 7.1 Mbps using UDP. IEEE 802.11g was the third modulation standard for Wireless LAN. It operates at a maximum raw data rate of 54 Mbps, or about 19 Mbps net throughputs (identical to 802.11a core, except for some additional legacy overhead for backward compatibility). The modulation scheme used in 802.11g is OFDM at the data rates of 6, 9, 12, 18, 24, 36, 48, and 54 Mbps.

1.2. Problem introduction

Of all the technologies which enabling the tremendous advances in data and voice communications, perhaps the most revolutionary is the development of cellular concept [5]. The use of cellular technology overcomes the capacity bottleneck when a limited spectrum is used for transmission – it provides system capacity expansion through the reuse of frequency over the geographical space. In cellular networks such as GSM, the service area is divided into many regions known as "cells". Each cell is allocated with a few frequencies and is served by a base station which consists of transmitters, receivers and a control unit. Adjacent cells are assigned with different frequencies to avoid interference or crosstalk. However, cells which are sufficiently separated can reuse the same frequency band for transmission. Other key methods used to improve the design of cellular networks are cell splitting, cell sectoring and the use of micro-cells, etc.

Wireless LANs provide mobility through roaming capabilities. However, because of the difference in multiple access signaling techniques used, deploying a wireless network is not simply a matter of identifying user locations and connecting them to the backbone. The deployment of each wireless system is unique in many aspects, and careful planning and a meticulous site survey are required. In the literature, the deployment of wireless networks involves the search for the base stations (BS) (or access points (APs)) placement while maximizing the profit or minimizing the cost has been studied. For this thesis, we narrow down our discussions to AP placement. However, the approach can be applied equally well to cellular systems. Normally, to minimize the cost, service providers (SPs) select the least number of APs to support the demand traffic, by taking both the multiple access and signaling techniques into consideration.

The solution to select a subset of candidate AP locations which is the least in number but yet able to cover all traffic demand points (DPs) is a combinatorial optimization problem – the well known minimum cardinality set covering problem [6]. The problem is NP-hard and heuristic approaches are usually used to obtain the suboptimal solutions. In [7], a computer-based tool which allows one to measure the AP coverage is developed. In [8], an efficient heuristic approach using a combined greedy and local search algorithm to reduce the computation time is reported. In [9], a framework based on simulated annealing is used for BS site selection. Rodrigues [10] gave out a mixed integer programming model to solve the wireless network deployment problem. Their objective is to maximize the sum of received signal power of all the mobile stations in the network. They also started to solve this problem using a commercial linear programming tool — CPLEX. Lee [11] proposed a MILP model to solve the deployment problem. Their optimization objective is to minimize the maximum channel utilization, which qualitatively is as indication of the user of congestion at the hottest spot in the WLAN service areas. In this paper, a method to dynamically adjust the configuration of the network to achieve its objective was mentioned. In [12], a method for finding optimal base stations configuration for CDMA systems jointly with uplink and downlink constraints was proposed using the approach.

In addition, the optimization problem can become quite complicated when multiple performance metrics need to be satisfied concurrently. Most of the reported works formulate the problem using mixed integer linear programming (MILP) but with different objective functions. In [8], the effect of interference was considered and results in a quadratic set covering problem. In [9], the objective function is to maximize the total received signal level of all the traffic DPs. In [10], the authors proposed an approach of optimizing AP placement to maximize the radio resource utilization. Normally these works assumed that the demand traffic does not vary with time. An exception is in [12], an integer linear programming (ILP) model is proposed and during the optimization process, the SP needs to observe the traffic demand over different periods of a day to obtain a better decision on AP sites.

1.3. Thesis motivation

The use of MILP model to solve such network deployment problems has been widely adopted. However, to my best knowledge, there is no effort to look into the deployment when multi-period optimization is of concern. If the business plan of SP is to look at return of investment, the placement of APs should adopt an intermittent approach: to match with the predicted future demand traffic over different periods of time rather than one solution for all. For example, as shown in Figure 1-1, in an indoor environment, a wireless network is to be deployed. There are totally 3 possible candidate AP locations which are labeled as A, B, and C. There are also 3 demand points which are labeled as 1,2 and 3. The cost of mounting an AP on location A,B and C are 10, 15 and 20 dollars separately due to different position and cable wiring. The circle around each candidate points indicates the effective transmission range of an AP if it is mounted on this point. In this picture, obviously, candidate A should be selected because the cost is the lowest (20 dollars) and all the demand points can be covered. However, if after some time, there are new demand points appear to be served, the planner will have to reconsider the problem. As is shown in Figure 1-2, there are 2 new demand points---demand points 4 and 5. The AP mounted on location A cannot cover demand point 4and 5. In this new scenario, an AP has to be mounted on candidate point C so that demand point 4 and 5 can be covered. Besides, once AP in C location is mounted, the AP in location A should be removed to reduce unnecessary cost. However, even if AP in location A is removed, some relevant cost such as mounting cost and wiring cost has already occurred in location A. If the AP can be reused, the total cost of mounting and wiring is the sum of mounting cost in A and C (30 dollars). What if we mount the AP in location C in the first place? The total cost should be 20 dollars only. So the optimal solution in one stage may not be the overall optimal solution over multiple time periods. If we can anticipate the future demand point location, maybe there is a way for us to improve the deployment plan.

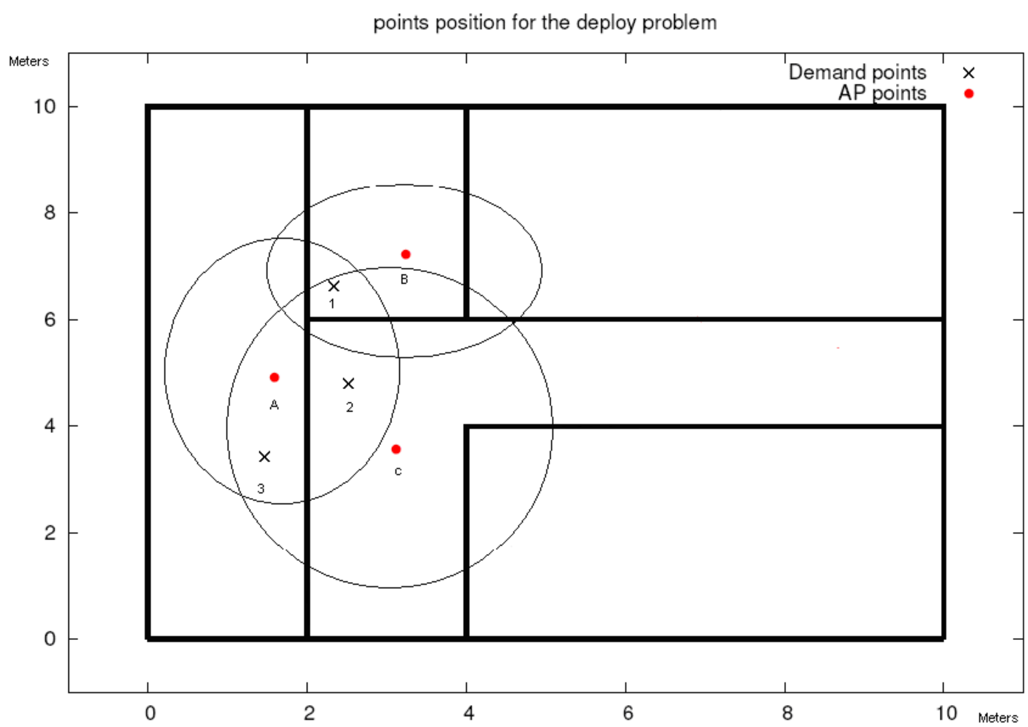


Figure 1-1 Candidate points and demand points in a indoor environment

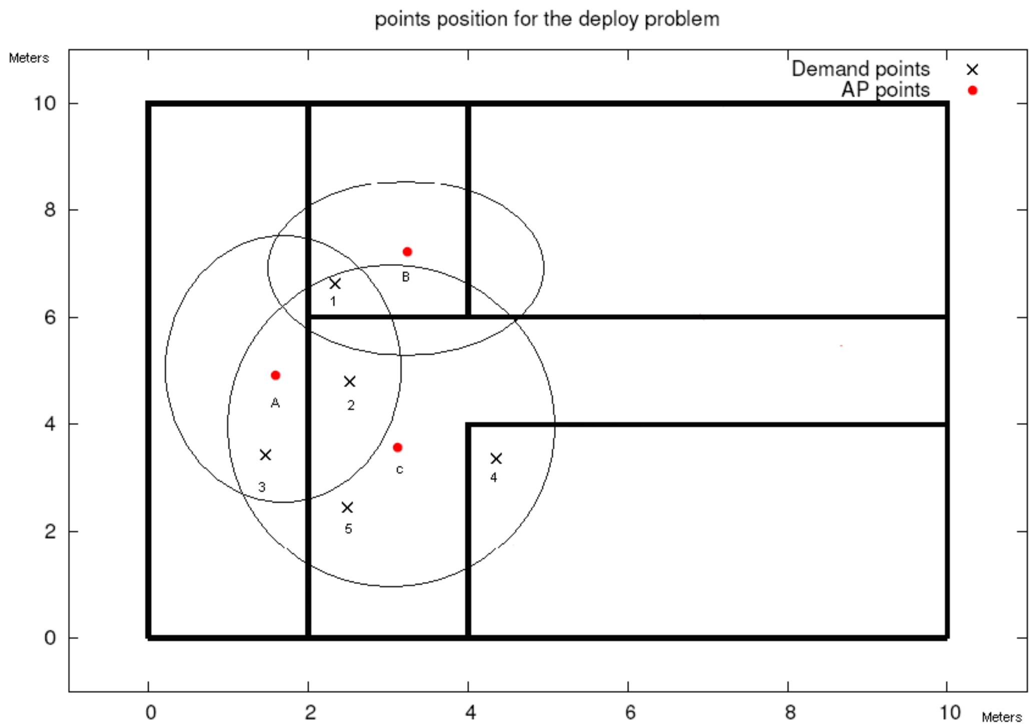


Figure 1-2 New demand points in indoor environment

This scenario is not rare in the real life, especially when certain network service is just taking off, the initial traffic can be low in each DP. As time goes by, the traffic in each DP will increase due to the acceptance of the technology by users. After some time, due to market saturation, the demand may maintain at a certain level. Figure 1-3 shows the change of the number of demand points in certain service area in different months. Because of this reason, service providers will need to find a solution to get as much profit as possible while keeping the customer satisfied with the service. The challenge is in the uncertainty of the future. This thesis developed a decision analysis model to solve this problem. This thesis also developed a probabilistic model for the scenario in indeterminate future.

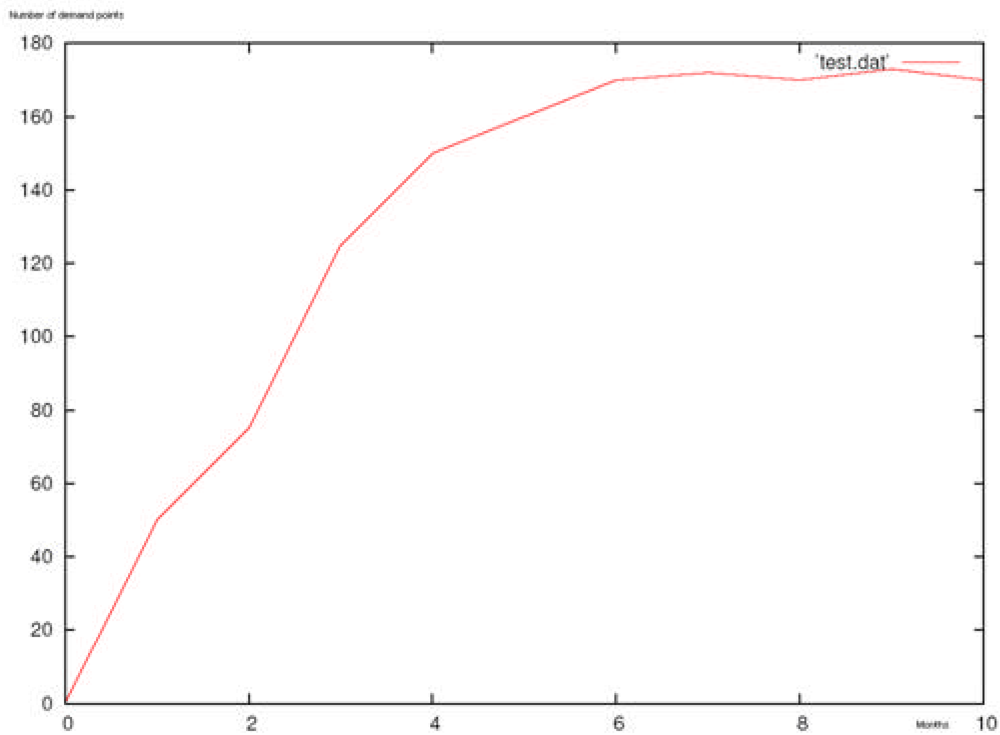


Figure 1-3 A projected traffic demand (versus time) at a given AP

1.4. Organization of the thesis

This thesis uses the method of operational research to deal with the network optimization. Our unique contribution is that optimization is performed over a longer time frame which includes the future projected traffic, hence, the solution may not necessary be optimized at the instant of the design.

In Chapter 2 and 3, two networks deploying different multiple access schemes are studied. Chapter 2 focuses on the FDMA system and Chapter 3 focuses on a WiFi-like system which deployed CSMA/CA. In Chapter 3, the effect of CSMA/CA protocol on the achievable throughput is discussed in detail. The estimation of the throughput is presented before brought into the integer programming model. Rate adaptation is also taken care in the model for the WiFi-like network. Chapter 4 focuses on the optimization over multiple periods for the wireless network. A branch-and-cut algorithm is introduced to improve the calculation speed. In Chapter 5, the decision analysis theory is used to deal with the uncertain future demand. Utility theory is also used when we try to adapt the different strategies to different users. Chapter 6 gives out the concluding remarks and directions for future research.

CHAPTER 2

SINGLE-PERIOD OPTIMIZATION FOR FDMA-BASED SYSTEMS

After we briefly introduced the optimization problem, we are going to give out our optimization model for the wireless networks. Let's start with the normally seen FDMA modulated network. FDMA is widely used in many wireless networks. In the traditional GSM systems, FDMA together with TDMA are used as the basic channel multiple access technique to separate users. In WLAN, FDMA is also used in 802.11b to avoid two nearby APs from using the same frequency for transmission. Significant research efforts have been made to optimize the cell coverage of wireless networks according to the demand traffic.

Design Problem

We consider the cell coverage problem for a FDMA system. We assume that at each AP, only one frequency channel could be allocated. This is normally the case in some wireless networks such as WLAN.

We assume the following inputs are provided in the design:

(a) Traffic demand in the service area

The service area is divided into demand points (DPs) known as grids, each with known demand traffic. Service providers (SPs) are assumed to be able to carryout study to find out the demand traffic at each AP.

(b) Candidate locations for APs

To solve for the optimal placement of APs over the service area has high complexity. To reduce the complexity, designers can pre-select sufficient candidate locations for the placement of APs. Other than complexity reduction, this approach is more practical since the locations suitable to place APs are normally constrained by physical factors such as walls, buildings and availability of power supply.

(c) Propagation model

Without loss in generality, we use a simple path loss and propagation model. If the losses resulting from floors and walls need to take into account, the model is given by

$$L = 40 + 20 \cdot \log(d) + n \cdot f + m \cdot w \quad ,$$

Where L is the pathloss in dB, n and m are the number of floors and walls between the transmit and receive antennas, f and w are the amount of attenuation (in dB) due to the presence of floors and walls in the propagation path. The path loss constant is set at 20 dB/decade for indoor applications. Typical attenuations for different obstacles are given in Table 2.1.

For properly designed networks, the transmission powers of APs after removing the effect of path loss/attenuations should be greater than the receiver sensitivity. In practice, a safe margin of a few dB should be given. Other parameters are: antenna gain = 3dBi; cable/ connector loss = -15dB; receiver sensitivity = -78dBm; margin = 15dB. Maximum transmission power is set at 31.4mW, which corresponds to a maximum coverage radius of 5 meters if no obstacle is presence.

Since the distances between the candidate APs and the DPs are known, the "additional" distances to account for the presence of obstacles can be calculated. For example, whenever there is an attenuation of mw dB due to the presence of walls, for the ease of deciding whether a DP at a physical distance d away is within the coverage of the AP, we define it as the equivalent distance D ($D = d + \Delta d$) which is given by

$$L = 40 + 20 \cdot \log(d \cdot 10^{mw/20}) = 40 + 20 \log D .$$

We make use of the fact that if a DP is serviced by an AP, then D must be within the maximum coverage radius $R_{\max}$ of the AP.

Obstacle	Attenuation [dB]
Floor	30
Brick wall with window	2
Office wall	6
Metal door in office wall	6
Cinder block wall	4
Metal door in brick wall	12.4
Brick wall next to metal door	3

Table 2.1 Typical value of attenuation for different obstacle

(d) Cost and revenue

To obtain the profit, the running costs need to be carefully considered. Investment costs include installation and equipment costs of APs which is a one-time expenditure. In the mean time, maintenance, power consumption, salaries and rental costs, which are normally proportional to the operating period. The revenue mainly comes from service charges or subscription fees.

Problem formulation

The purpose of the network deployment problem is to obtain the installation plan which maximizes the profit of an investment over a longer period of time. This includes the intermittent placement of APs at different periods by jointly considering the probabilistic or deterministic future demand traffic.

Our first step is to solve the optimal solution at a given period of time for the given demand traffic. The problem is formulated using a MILP model. Suppose there is a set of N candidate sites $A = \{1, \cdots, N\}$ for APs, and a set M DPs $D = \{1, \cdots, M\}$. Each DP needs to select an AP from the active APs while fulfilling the operation constraints. The demand traffic at DP j is denoted by T_j and i^{th} AP supports a portion of this traffic t_{ij}.

Let x_{ij} be the assignment parameter denoting the link between DP j and candidate AP site i. $x_{ij} = 0$ if there is no link between j and i, $x_{ij} = 1$ if AP i is to support the demand traffic at i. Let a_i is used to indicate whether AP i should be activated. $a_i = 0$ if the i^{th} AP is not activated, $a_i = 1$ if the AP is activated. Because a link exists only if that particular AP is being selected,

$$x_{ij} \le a_i \quad \forall i \in A \quad \forall j \in D \tag{2.1}$$

Let c_{ik} be the assignment parameter indicating whether channel k of AP i is being activated. The channel k must be selected from a K predefined channels each having an operating frequency and capacity B. Suppose t_{ij} is the traffic demand from DP j to AP i, then

$$\sum_{j=1}^{M} t_{ij} \le \sum_{k=1}^{K} c_{ik} \cdot B \quad \forall i \in A \tag{2.2}$$

This means that the total traffic supported by an AP cannot exceed its own capacity. In the mean time, only the activated AP can have allocated channels, so

$$\sum_{k=1}^{K} c_{ik} \le K \cdot a_i \quad \forall i \in A \tag{2.3}$$

In case if only one frequency is to be used at each AP, (3.3) can be written as $\sum_{k=1}^{K} c_{ik} \le a_i$.

For each DP, the sum of the potential traffic supported by all the activated APs should exceed its traffic demand, i.e.,

$$\sum_{i=1}^{N} t_{ij} \ge g \cdot T_j \quad \forall j \in D \tag{2.4}$$

where $g \geq 1$ is a factor to reserve some bandwidth in the APs. Meanwhile, since only those activated APs can support traffic, hence

$$t_{ij} \leq g \cdot x_{ij} \cdot T_j \qquad \forall i \in A \quad \forall j \in D \tag{2.5}$$

Let r_i denote the equivalent coverage radius (i.e., attenuation due to the presence of obstacles has been compensated by an additional distance) of the AP for its transmission power. Let D_{ij} denote the equivalent distance between AP i and DP j. If there is a link between i and j, then the radius of the AP i must be larger than the calculated equivalent distance between the DP and the AP itself.

$$x_{ij} \cdot D_{ij} \leq r_i \qquad \forall i \in A \quad \forall j \in D \tag{2.6}$$

Obviously, the coverage of AP i_1 and AP i_2, denoted by r_{i_1} and r_{i_2}, respectively, should satisfy the relationship $r_{i_1} + r_{i_2} \leq D_{i_1 i_2}$ for the same frequency to be reused without interfering each other. We define a binary variable $f_{i_1 i_2}$ and $f_{i_1 i_2} = 0$ is used to indicate no interfere between APs i_1 and i_2. The following relationship is used to represent such requirement:

$$r_{i_1} + r_{i_2} \leq D_{i_1 i_2} + f_{i_1 i_2} \cdot (2R_{\max} - D_{i_1 i_2}) \quad \forall i_1, i_2 \in A \tag{2.7}$$

In (3.7), non-interfere condition fulfils only if $f_{i_1 i_2} = 0$. For $f_{i_1 i_2} = 1$, the relationship $r_{i_1} + r_{i_2} \leq 2R_{\max}$ gives no effect on the solution since by definition the inequality always holds.

For FDMA system, different frequencies are used for different channels. To avoid interference between neighboring APs, we introduce the constraint:

$$c_{i_1 k} + c_{i_2 k} \leq 2 - f_{i_1 i_2} \qquad \forall i_1, i_2 \in A, \ \forall k \tag{2.8}$$

Sometimes we would like to impose a limit on the excess bandwidth each AP can provide to prevent resource wastage. Let $k \geq g \geq 1$ denote this predefined factor, then

$$\sum_{i=1}^{N} t_{ij} \leq k \cdot T_j \qquad \forall i \in A \quad \forall j \in D \tag{2.9}$$

Other constraints to be included are:

$t_{ij} \geq 0$; $r_{ij} > 0$; $c_{ik}, x_{ij}, a_i \in \{0,1\}$.

Let P denote the utilization charge for per unit demand traffic, C_A is the hardware cost, $C_{I\,i}$ denote the initialization cost such as installation, $C_{M\,i}$ denote the maintenance cost of AP i and C_c denote the running cost of a channel. So we have the objective function as follows:

$$\max \quad R = \sum_{i=1}^{N} T_i \cdot P - \sum_{i=1}^{N} a_i (C_{M_i} + C_{I_i} + C_A) - \sum_{i=1}^{N} \sum_{k=1}^{K} c_{ik} \cdot C_c \tag{2.10}$$

The design has now formulated into a MILP problem. Note that in the objective function, the revenue is a constant and do not involve in the optimization process, hence we exclude it and expect a negative value on the profit. The term $\sum_{i=1}^{N} \sum_{j=1}^{M} t_{ij} - \sum_{i=1}^{N} T_i$ accounts for excess bandwidth reserved to ensure the overall system grade-of-service and we do not introduce any penalty to the objection function. Further we assume the capacity per network card is a constant and hence the excess bandwidth will be zero. Another assumption is bandwidth is available for transmission in a scheduled manner, i.e., the time for terminals to contend for bandwidth is negligible. For WLAN where CSMA/CA protocol is used, the approximate approach can be found in Chapter 3.

Optimal Solution

We take the following network deployment problem as our example. We assume that at each AP, only one channel will be used.

Example 2.1: We use the layout of a service area shown in Figure 2-1.

The lines drawn in black represent walls; each black marker represents a traffic DP and each red marker represents a candidate site for APs. There are 8 candidate sites for APs and 15 DPs in this office. We assume that each AP uses only one frequency. The values of parameters used are: $B = 11$, $K = 3$, ? =1.2, ? =1.1, $R_{\max}$ = 5.00, $C_C = 5$. $C_m = 15$ and $C_A = 20$ are the same for all AP. The values of $C_{l\,i}$ for each candidate AP sites together with the co-ordinates are given in Table 2.2. The demand of traffic and the co-ordinates can be found in Table 2.3.

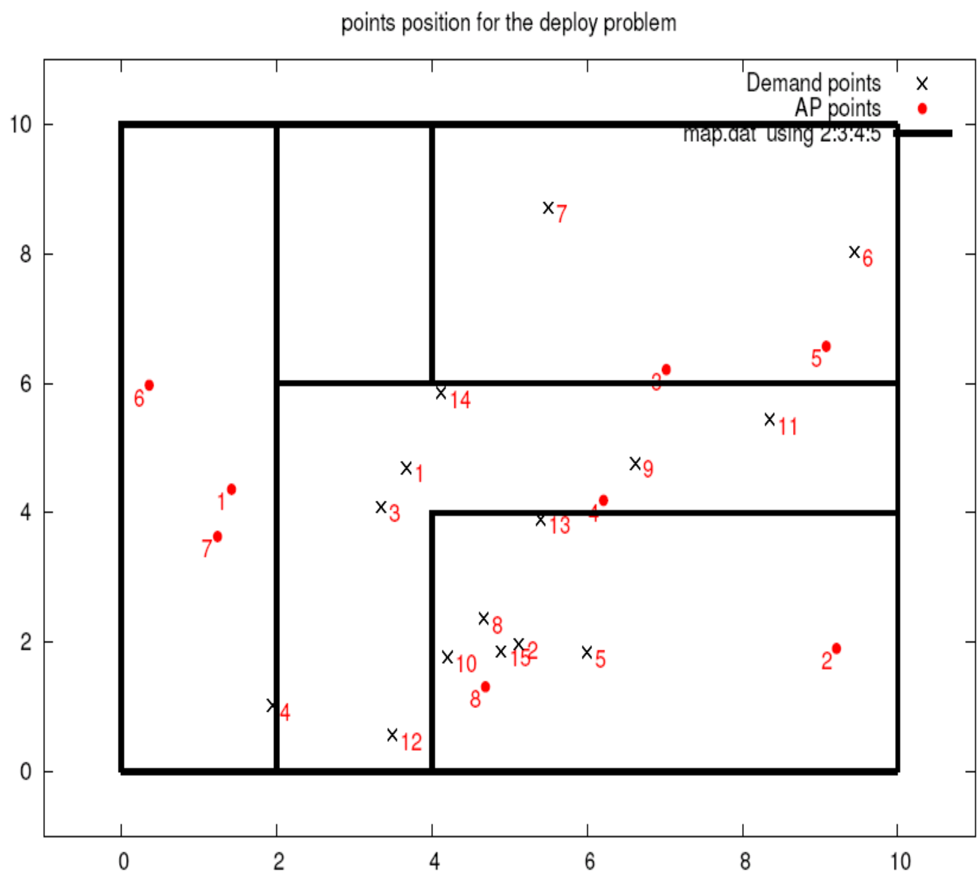


Figure 2-1 An example showing the traffic demand points (crosses) and candidate access point sites(points)

AP	X	Y	Initial cost	AP	X	Y	Initial cost
1	1.42	4.36	142.98	2	9.21	1.90	124.25
3	7.02	6.21	137.35	4	6.21	4.19	146.35
5	9.08	6.57	122.02	6	0.36	5.97	107.27
7	1.24	3.63	133.89	8	4.69	1.31	116.27

Table 2.2 The initial costs for each candidate APs (C_i)

DP	X	Y	Demand	DP	X(m)	Y(m)	Demand
1	3.67	4.69	2.13	2	5.12	1.97	2.75
3	3.35	4.09	3.1	4	1.94	1.02	2.68
5	6.00	1.84	2.62	6	9.44	8.03	2.2
7	5.50	8.71	2.15	8	4.67	2.36	2.86
9	6.62	4.76	2.03	10	4.20	1.77	4.5
11	8.35	5.44	2.34	12	3.49	0.56	2.97
13	5.40	3.90	2.1	14	4.12	5.85	2.51
15	4.89	1.85	2.56				

Table 2.3 Demand traffic at each demand point (T_j)

The solution is also labeled in Figure 2-2. It can be seen that 3 APs are selected. They are candidate points 3, 7 and 8. We can also get the transmission radius of the three selected APs. $r_3 = 4.92$ m, $r_7 = 4.65$ m and $r_8 = 3.42$ m. The path loss from the AP to the edge of its transmission radius can be calculated to be $L_3 = 53.84$ dB, $L_7 = 53.35$ dB $L_8 = 50.68$ dB. The transmission powers are $P_3 = 30.47$, $P_7 = 27.22$ and $P_8 = 14.72$ mW, respectively. GNU GLPK programming toolkit is used as the linear programming engine. The whole calculation process takes about 7 minutes to finish. From this result, we can see that the linear programming optimization can optimize the resource allocation. Each AP tries to serve as many demand points as possible.

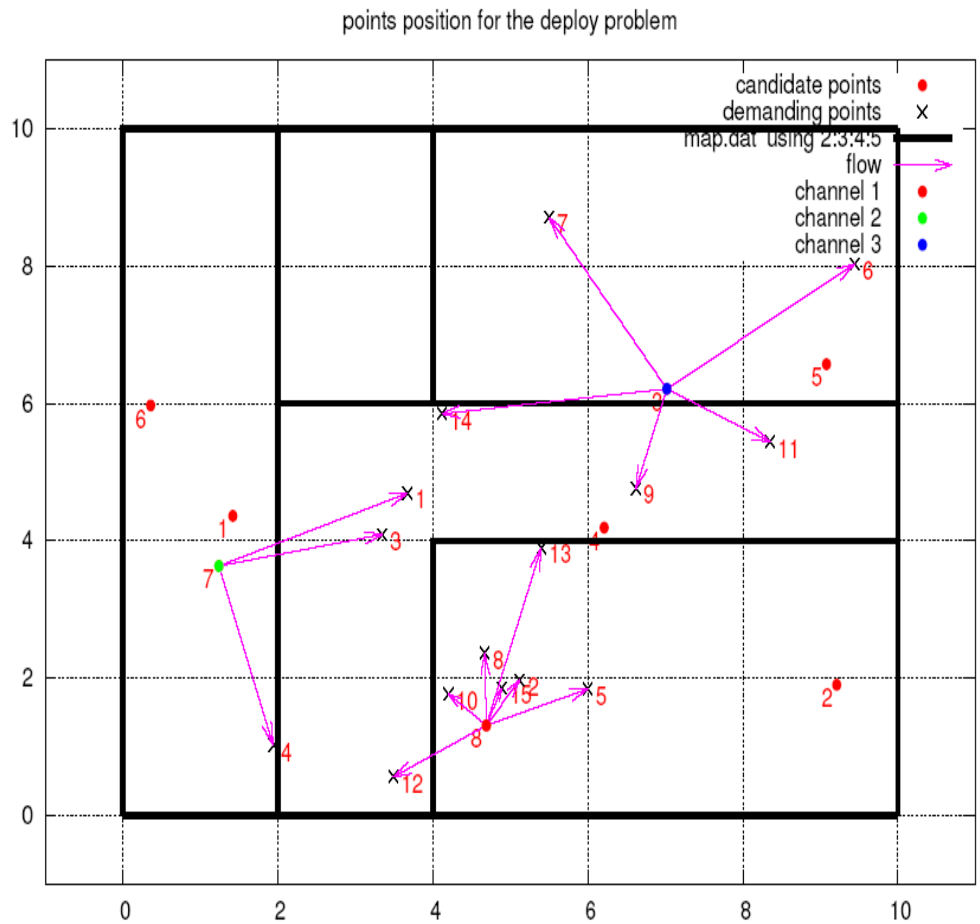


Figure 2-2 Optimal AP placements to support the given traffic demands

(DPs are marked in crosses and candidate AP sites are marked in dots)

CHAPTER 3

SINGLE-PERIOD OPTIMIZATION FOR CSMA-BASED SYSTEMS

A WiFi-like wireless network shares many attributes of the traditional FDMA wireless network as there are several frequency bands for it to operate on. However, it is unique in two important aspects: the use of CSMA/CA has impact on the achievable throughput and the use of rate adaptation to handle different signal qualities.

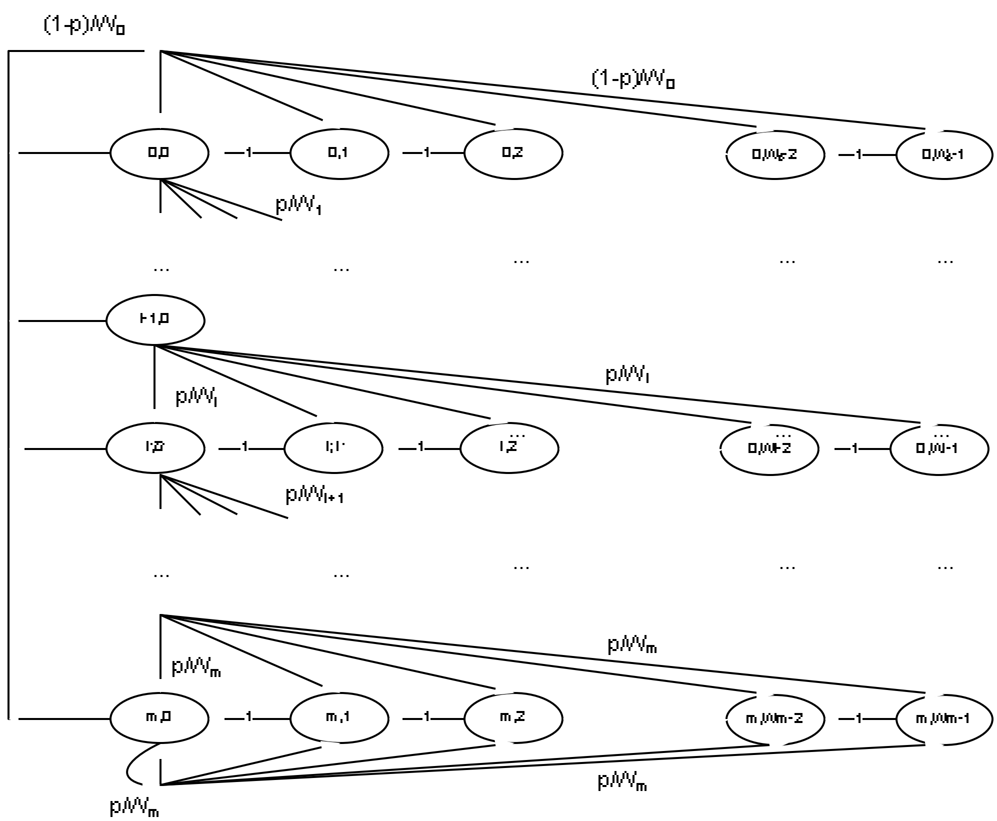
3.1 Throughput of CSMA/CA based systems

Significant amount of effort has been performed to study the available throughput when CSMA/CA is used. In [13], the effect of CSMA/CA protocol is analysed for voice users. The result shows that the throughput of the system cannot achieve its theoretical maximum due to the contention. In [14], a theoretical analysis is made to estimate the performance using CSMA/CA protocol. A state machine was proposed to formulate the problem. The result also shows that the throughput will enter a saturated condition when the number of mobile stations increases to a certain value. In [15], capture probability in fading and shadowing channels is considered. The result shows that capture probability converges to a finite limit as the number of colliding terminals increases. Furthermore, the authors have developed a new analytical approximation for the performance of CSMA/CA protocols in the presence of Rayleigh fading, lognormal shadowing, and power capture effect. They have found that the saturated throughput of CSMA/CA protocols remains the same as the number of terminals and the offered load increases. The throughput performance of CSMA/CA protocols can also be affected by the average packet size in transmission. In [16], the performance under different packet sizes is studied, the authors presented the exact formulae for the throughput of IEEE 802.11 networks with no transmission error under various physical layers, data rates and packet sizes. By performing this calculation, the authors concluded that the theoretical maximum throughput of a WLAN channel is sensitive to the packet size. The author found that the higher the data rate, the more sensitive the packet size to the maximum throughput. Performance will be substantially degraded when small-sized data packets are transmitted at high data rates. In [17], the paper presented an exact analysis for the performance of wireless LANs which support real-time and data communication using CSMA/CA protocol. They concluded that the network performance can be improved by adequately choosing the ratio between the collision avoidance period and the transmission period. In [18], the paper proposed a method that can evaluate the cell throughput performance of wireless local area network (LAN) systems in which the CSMA/CA and multiple transmission bit-rate (multi-rate) techniques are used. The analysis confirmed that the approach of expanding the bandwidth of the IEEE 802.11a system is effective to improve the cell throughput performance. The authors in [19] provided a theoretical framework to optimize single-user throughput by selecting the transmitted bit rate and payload size as a function of channel conditions, for both AWGN and Nakagami-*m* fading channels. Numerical results revealed that careful payload adaptation can significantly improve the throughput performance at low SNRs, whilst at higher SNRs, rate adaptation with higher payload length provides better performance. In [20] and [21], the analysis of the CSMA/CA protocol is made, where a throughput model based on Markovian analysis is proposed for CSMA/CA networks with a general topology. Simulation investigations are presented to verify its performance. Fairness issues in CSMA/CA networks are also discussed.

Our earlier analysis in Section 2.1 does not include CSMA protocol and SNR issue and hence is more applicable to a centralized model where DP's traffic will be instructed by the centralized controller to transmit. In some applications where APs are not coordinated, the effect of CSMA/CA protocol cannot be ignored. We now try to integrate this protocol into our earlier integer programming model. The primary MAC technique of 802.11 is called distributed coordination function (DCF). DCF is a CSMA/CA scheme with binary slotted exponential backoff. DCF describes two techniques to be employed for packet transmission. The default scheme is a two-way handshaking technique called Basic Access mechanism. This mechanism is characterized by the immediate transmission of a positive ACK by the destination station, upon successful reception of a packet transmitted by the sender station. Explicit transmission of an ACK is required since, in the wireless medium, a transmitter cannot determine if a packet is successfully received by listening to its own transmission. In addition to the Basic Access, an optional four way handshaking technique, known as Request-To-Send/Clear-To-Send (RTS/CTS) mechanism has been standardized. Before transmitting a packet, a station operating in RTS/CTS mode “reserves” the channel by sending a special RTS message. The destination station acknowledges the receipt of a CTS frame by sending back a CTS frame, after which normal packet transmission and ACK response occurs [4]. Since collision may occur only on the RTS frame, and it is detected by the lack of CTS response, the RTS/CTS mechanism allows increasing the system performance by reducing the duration of a collision when long messages are transmitted.

DCF adopts an exponential backoff scheme. At each packet transmission, the backoff time is uniformly chosen in the range $(0, w-1)$. The value w is called contention window, and depends on the number of transmissions failed for the packet. At the first transmission attempt, w is set equal to a value $W_{\min}$ called minimum contention window. After each unsuccessful transmission, w is doubled, up to a maximum value $W_{\max} = 2^m W_{\min}$. The backoff time counter is decremented as long as the channel is sensed idle, remains unchanged when a transmission is detected on the channel, and re-activated when the channel is sensed idle again for more than a DCF Inter frame Space (DIFS). The station transmits when the backoff time reaches 0.

In [20], a Markov chain model for the backoff window size is proposed with the assumption that at each transmission attempt, and regardless of the number of retransmissions suffered, each packet collides with constant and independent probability p. Let $s(t)$ be the stochastic process representing the backoff stage $(0,\ldots m)$ of the station at time t. Let $b(t)$ be the stochastic process representing the backoff time counter for a given station. A bi-dimensional process $\{(s(t),b(t)\}$ with the discrete-time Markov chain can be depicted in the Figure 3-1.



In this Markov chain, the transition probabilities are

$$\begin{cases} P\{i,k \mid i,k+1\} = 1 & k \in (0, W_i - 2), i \in (0, m) \\ P\{0,k \mid i,0\} = (1-p)/W_0 & k \in (0, W_0 - 1), i \in (0, m) \\ P\{i,k \mid i-1,0\} = p/W_i & k \in (0, W_i - 1), i \in (1, m) \\ P\{m,k \mid m,0\} = p/W_m & k \in (0, W_m - 1) \end{cases} \tag{3.1}$$

The first equation in (4.1) accounts for the fact that, at the beginning of each slot time, the backoff time is decremented. The second equation accounts for the fact that a new packet following a successful packet transmission starts with backoff stage 0, and thus the backoff is initially uniformly chosen in the range $(0, W_0 - 1)$. The other two cases model the system after an unsuccessful transmission. In particular, in the third equation, when an unsuccessful transmission occurs at backoff stage $i-1$, the backoff stage increases, and the new initial backoff value is uniformly chosen in the range $(0, w_i)$. Finally, the fourth equation models the fact that once the backoff stage reaches the value m, it does not increase in subsequent packet transmissions.

In [20], the probability of a station transmitting in a randomly chosen slot time t in terms of p is

$$t(p) = \frac{2}{1 + W_{\min} + pW_{\min} \sum_{i=0}^{m-1} (2p)^i} \tag{3.2}$$

Let S be the normalized system throughput, defined as the fraction of time the channel is used to successfully transmit payload bits. To compute S, let us analyze what can happen in a randomly chosen slot time. Let P_{tr} be the probability that there is at least one transmission in the considered slot time. Since n stations contend on the channel, and each transmits with probability

$$P_{tr} = 1 - (1 - t)^n \tag{3.3}$$

The probability P_s that a transmission occurring on the channel is successful is given by the probability that exactly one station transmits on the channel, conditioned on the fact that at least one station transmits, i.e.:

$$P_s = \frac{nt(1-t)^{n-1}}{1-(1-t)^n} \tag{3.4}$$

S can then be expressed as follows:

$$S = \frac{P_s P_{tr} E[L]}{(1-P_{tr})s + P_{tr} P_s T_s + P_{tr}(1-P_s) T_c} \tag{3.5}$$

Here, T_s is the average time the channel is sensed busy (i.e.the slot time lasts) because of a successful transmission, and T_c is the average time the channel is sensed busy by each station during a collision. s is the duration of an empty slot time. L is the payload size. $E[L]$ is the average packet payload size.

In [20], a saturation throughput is calculated as n grows to a sufficiently large number. The maximum achievable throughput $S_{\max}$ can be approximated as:

$$S_{\max} = \frac{E[L]}{T_s + sK + T_c(K(e^{1/K}-1)-1)} \tag{3.6}$$

Here, K is a constant (takes a value 9.334 for the Basic Access mechanism and 2.042 for the RTS/CTS mechanism) and is irrelevant to the number of mobile stations, n [20]. According to the author's calculation, the saturation throughput for Basic Access is about 0.823957 and 0.835859 for the RTS/CTS mechanism.

However, before the saturation throughput is reached, the throughput depends on several factors, which are summarized in the following Table 3.1.

Parameters	Basic Access	RTS/CTS
Network size	Sensitive	Insensitive
Prob. t $W_{\min}$	Sensitive Dependent	Insensitive Independent
$W_{\max}$	Marginal effect	Negligible effect
Packet size	Less effective for longer packets than RTS/CTS	More effective for longer packets

Table 3.1 Parameters that affects throughput

3.2 The distance effect on throughput

Besides the effect of CSMA/CA protocol, there are other factors that can affect the throughput of WiFi-like network. One of these is the distance effect. Ref [22] reported on a study that looked at the performance of an 802.11b WLAN for a small user community. The author concluded that the performance of the WLAN is relatively not affected by WEP (wired equivalent privacy) encryption, number of concurrent users, or offered load, however, the increased distance of the user from the wireless access point results in degraded performance. Abrahams [23] analyzed the relationship between the SNR and the throughput. His conclusion is that the throughput can be affected by the distance because of the SNR requirement for the received signal as a result of rate adaptation. In [24], a study of the Orinoco PCMCIA Silver/Gold WLAN card shows the SNR relationships with throughput are given by: 11Mbps → 16 dB; 5.5 Mbps → 11 dB; 2 Mbps → 7 dB; 1 Mbps → 4 dB.

According to this fact, wireless networks such as WLAN which employed rate adaptation, the mobile stations within the coverage of an AP may not be transmitting at the same data rate. When the mobiles are closer to AP, higher data transmission rate could be achieved.

According to the exponential path loss model, the received powers at two locations d_1 and d_2 away from an AP, denoted by P_1 and P_2 respectively, have the following relationship:

$$\frac{P_1}{P_2} = \left(\frac{d_1}{d_2}\right)^{-n} \tag{3.7}$$

For an AP with fixed transmission power, the maximum transmission range is also fixed. If we define the maximum coverage radius for an AP as d_{max}, and the corresponding minimum transmission power to achieve this maximum transmission range as $P_{min,}$ for any mobile that is within the AP's transmission range, we have

$$P = \left(\frac{d}{d_{max}}\right)^{-n} \times P_{min}. \tag{3.8}$$

Here P denotes the received power at the DP, n is the path loss exponent, which typically takes a value of 2 for indoor WLAN. Using (4.8), we can deduce the ratio between d and d_{max} which can ensure that P is greater than a desired received power threshold to achieve a given transmission rate. Hereafter, we define $g = \left(\frac{d}{d_{max}}\right)$.

The values of g defined the regions where throughput can be achieved under a given n. We illustrate this using an example. Given the minimum received power threshold at the boarder of the AP's transmission range is -94dBm (e.g, Orinocco cards PCMCIA Silver/Gold), which can ensure 1 Mbps data rate. Other parameters are: antenna gain = 3dBi; cable/ connector loss = -15dB; margin = 15dB. Under these conditions, the required transmission power should be -94dBm + 15dB + 15dB - 3dB

= - 67dBm. Now we have the same mobile device which requires 2 Mbps data rate which will require -91 dBm receiver sensitivity. The required transmission power for this device can be calculated to be -64dBm. From (4.8) we can deduce that g equals to 0.841. It means that for n=2, at a distance 0.841 times away from the maximum coverage radius, 2 Mbps data rate can be achieved and beyond this distance the maximum achievable transmission rate can only be 1 Mbps. In Figure 4-2, the data rate fall off effect as a function of distance from an AP is shown. We assume the maximum radius of the AP is $d_{\max}$ where the achievable data rate is 1Mbps. Hence we calculated the g for the places where the achievable data rate can be 2, 5.5 and 11Mbps. The values of g for these 3 data rates should be 0.841, 0.668 and 0.501, respectively, for $n = 2$.

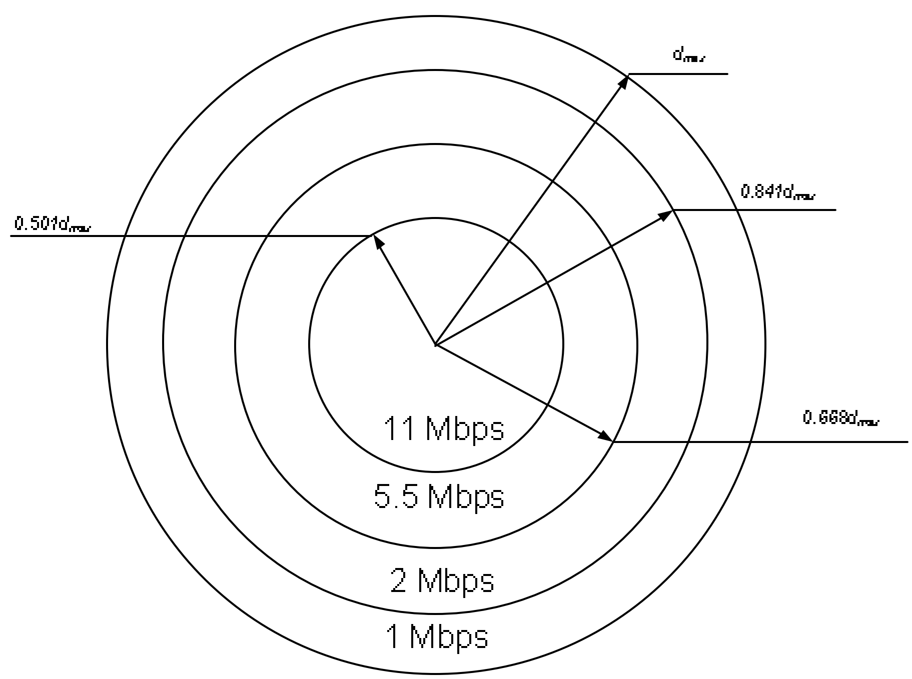


Figure 3-2 Data rate fall off for 802.11b as function of distance from AP for $n = 2$

data rate \ n	2	3	4
1 Mbps	1	1	1
2 Mbps	0.841395	0.891251	0.917276
5.5 Mbps	0.668344	0.764422	0.817523
11 Mbps	0.501187	0.630957	0.707946

Table 3.2 The value of g for different value of path loss n:

In our design, users are assumed to specify a suitable g according to the data rate requirements so that the desired data rate is taken into consideration in the optimization process.

3.2 Design for a WiFi-like network

The accurate cellular coverage design can be very difficult if we take the CSMA mechanism and rate adaptation into consideration. This is because of the factors such as throughput, transmission power and how DP is being distributed over the entire service area. These factors are intertwined with each other in the design process. However, an approximate approach similar to the FDMA model can be adopted by making some modifications to some of the constraints.

(a) Throughput limit based on CSMA/CA

As is analyzed in the literature, the throughput of CSMA/CA protocol has been computed and it is a constant value when the number of APs is sufficiently large.In our optimization model, we use the saturation throughput to perform the design. The MILP model is similar to the FDMA model. However, (2.2) is modified to:

$$\sum_{j=1}^{M} t_{ij} \le B_s \tag{3.9}$$

Here, B_s is the saturated throughput calculated based on the original bandwidth.

(b) Rate adaptation of the base stations

To guarantee the throughput for each mobile station, each DP can specify their quality-of-service in terms of desirable transmission rate. In our approach, this is achieved by making each DP specifying a value of g. We add new equations to the MILP model such as

$$x_{ij} \cdot D_{ij} \le g_j \cdot r_i \qquad \forall i \in A \quad \forall j \in D \tag{3.10}$$

Eq. (3.10) means that if a DP is connecting to an AP, the possible transmission rate depends on its distance from the AP. This equation ensures that the rate adaptation requirement at a given g_j can be guaranteed.

3.3 Optimal solution over a period

We take the following network deployment problem as our example. We assume that at each AP, only one channel will be used.

Example 3.1: We use the layout of a service area shown in Figure 2-2. The lines drawn in black represent walls; each black marker represents a traffic DP and each red marker represents a candidate site for APs. There are 8 candidate sites for APs and 15

DPs in this office. We assume that each AP uses only one frequency. The values of parameters used are: $B = 11$, $K = 3$, ? =1.2, $R_{max} =$ 50 meters, $C_C = 5$. $C_m = 15$ and $C_A = 20$ are the same for all AP. The values of $C_{I\,i}$ for each candidate AP sites together with the co-ordinates are given in Table 3.3. The demand of traffic and the co-ordinates can be found in Table 2.3.

AP	X	Y	Initial cost	AP	X	Y	Initial cost
1	53.9	32.8	109.77	5	57.2	33.2	126.07
2	6.8	98.3	118.64	6	55.3	76.8	134.86
3	45.1	2.2	102.71	7	88.9	59.9	147.03
4	10.7	39	129.64	8	95.1	79.8	107.44

Table 3.3 The initial costs for each candidate APs (C_I)

For the demand points, we have the following parameters: the location information (x and y), the total demand and the class of service for this point. The classes are divided into 3 groups with the class number 1 to 3, from the highest required throughput to the lowest. The corresponding ratios to the radius of the AP are 0.6, 0.8 and 1 separately.

We set the receiver sensibility to be -78dB, and the cable loss at both sides to be 10dB, antenna gain at both sides is 8 dB, we have 20 dB remaining margin for the receive side.

DP	X(m)	Y(m)	Total Demand	Class of service
1	56	88.4	0.43	3
2	31.6	28.2	2.79	2
3	92.9	90.8	2.17	2
4	77.6	82.1	1.73	3
5	48	29.6	4.43	3
6	59.2	39.5	0.57	3
7	72.8	4.2	2.23	2
8	33.3	13.1	2.4	1
9	90.7	82	4.51	1
10	60.3	11.7	4.3	1
11	10.2	42.9	1.64	2
12	84.6	58	2.65	1
13	9.2	68.9	0.52	3
14	30.2	87	3.17	2
15	62.9	68.9	2.91	1

Table 3.4 Demand traffic at each demand point (T_j)

The solutions are shown in Figure 3-3 and the computed values are tabulated in Table 3.5. From the result, we can see that if we take the rate adaptation into consideration, the cost will increase and the total profit will decline. If we set the maximum transmission range to be higher, we can see the two profits are getting closer. This is because with the APs' transmission radius increase, more and more demand points are falling into the feasible range of these APs. Figure 3-4 shows the different profit under different transmission ranges:

:

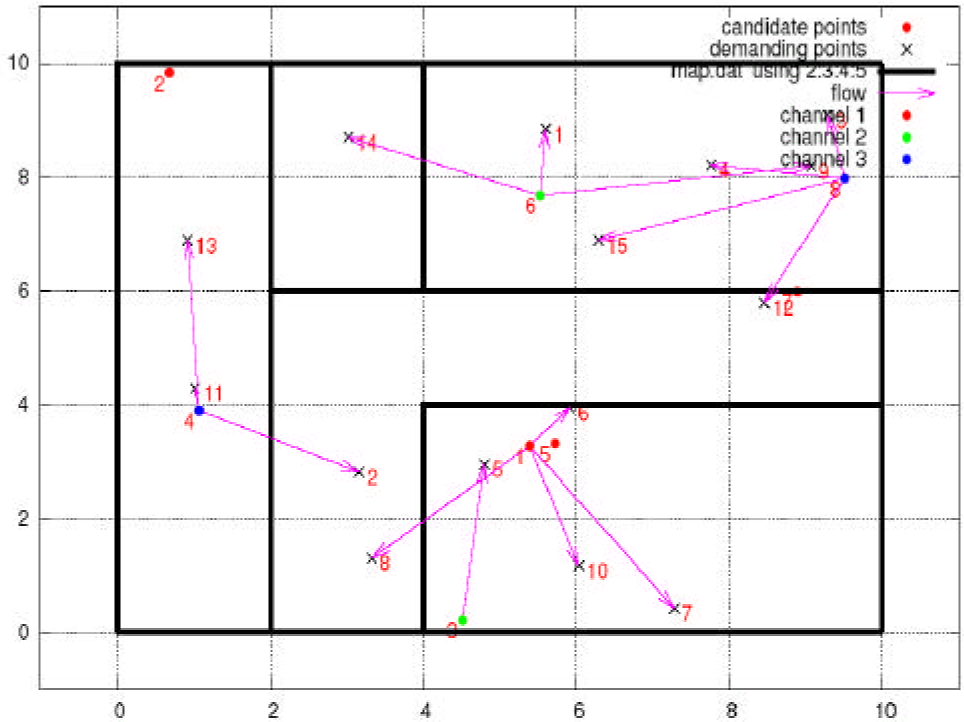


Figure 3-3 AP deployment without classes

AP	Solution with no rate adaptation			Solution with rate adaptation		
	selected	radius	Power(mw)	Selected	radius	Power(mw)
1	Yes	48.5	102.67	Yes	48.5	102.67
2	No	0		No	0	
3	Yes	27.6	33.25	Yes	28.2	34.71
4	Yes	43.5	82.6	Yes	43.5	82.6
5	No	0		No	0	
6	Yes	47.7	98.489	Yes	47.1	96.84
7	No	0		Yes	37.4	61.05
8	Yes	44.2	85.28	No	0	
profit	-369.92			-458.71		

Table 3.5 Results comparison for with and without rate adaptation

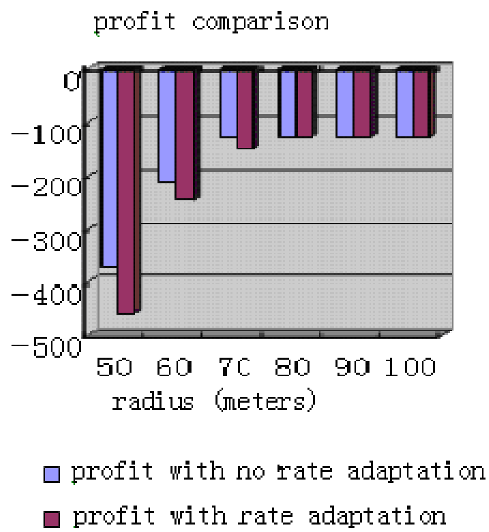


Figure 3-4 Comparison between with and without rate adaptation

CHAPTER 4

OPTIMIZATION MODEL CONSIDERING FUTURE TRAFFIC

From practical viewpoints, there is more factors still need to be taken into consideration. For example, the deployment of wireless networks needs to consider longer term investment return. The gradual increase of demand traffic imposes a challenge to the design: closing and re-opening of APs so as to match the changes in traffic demand over different periods of time. The solution for a long term profit is a challenging problem because optimized solution in each period may not necessary guarantee overall optimal. Besides, in practice, the prediction of future traffic cannot be accurately made as a result of competition from other SPs and the acceptability of services.

4.1 Formulation

We consider FDMA network with the assumption that in each period, the demand at each DP is a known value. As mentioned before, the traffic demand will vary after its deployment, especially at the technology taking off period; the change tends to be larger. This complicates the design as inappropriate placement of APs during the earlier stage may result in much larger cost in the future. For example, in Period I, by installing both AP1 and AP2 give the optimal profit whilst in Period II, AP3, AP4 and AP5 give the optimal. This may cause two APs to be relocated And a new AP is also needed in Period II. However, another solution which proposed to install AP3 and AP4 in Period I although is not optimal but may still give overall optimal if the loss in profit in Period I is smaller than the extra cost to carryout relocation of APs in Period II.

We now express the new objective function over multiple periods mathematically. For simplicity, we consider the case where each AP uses only one frequency set. The objective function at the p^{th} period is given by

$$\max \qquad R^{(p)} = -\sum_{i=1}^{N} a_i^{(p)} (C_M^{(p)} + C_{I_i} + C_A + C_C) \tag{4.1}$$

Hereafter, we use the superscript $^{(p)}$ to represent the parameter values at period p. $R^{(p)}$ represents the reverse of the total cost in period p.

For two consecutive periods p and $p+1$, it is not difficult to figure out that the number of APs which are common over the two periods is given by $n_C^{(p)} = \sum_{i=1}^{N} a_i^{(p)} a_i^{(p+1)}$ and the total number of distinct activated APs is given by

$n_d^{(p)} = \max \{\sum_{i=1}^{N} a_i^{(p)}, \sum_{i=1}^{N} a_i^{(p+1)}\}$. Hence, the number of APs which has to be relocated is $n_L^{(p)} = [\min\{\sum_{i=1}^{N} a_i^{(p)}, \sum_{i=1}^{N} a_i^{(p+1)}\} - n_C]$. The objective function is given by

$$
\begin{aligned}
\max \quad & R^{(p)} + R^{(p+1)} = \\
& -\sum_{i=1}^{N} a_i^{(p)} (C_M^{(p)} + C_{I_i} + C_C) - \sum_{i=1}^{N} a_i^{(p+1)} (C_M^{(p+1)} + C_{I_i} + C_C) \\
& + \sum_{i=1}^{N} a_i^{(p)} a_i^{(p+1)} C_{I_i} - n_d^{(p)} \cdot C_A
\end{aligned}
\tag{4.2}
$$

The third term in (4.2) is to remove the over-claimed costs due to the common APs used in the two periods. The fourth term is the total cost of hardware over the two periods. By recursively applying (4.2) to P periods, we have the objective function for the long term investment is given by

$$
\begin{aligned}
\max \quad R_T &= \frac{1}{2}\left[R^{(1)} + R^{(P)}\right] + \frac{1}{2}\sum_{p=1}^{P-1}\left[R^{(p)} + R^{(p+1)}\right] \\
&\leq \sum_{p=1}^{P} \left[\max R^{(p)}\right]
\end{aligned}
\tag{4.3}
$$

Eq. (4.2) can be substituted to obtain the objective function over multiple periods. The constraints listed from (2.1) to (2.9) are still applied to each period. Note that the objective function now will become nonlinear and the solution complexity becomes very high due to the large number of variables (equal to p times of the number of unknowns in each period).

4.2 Reduce the number of feasible solutions

We proposed a complexity reduction searching algorithm without sacrificing any accuracy. The approach is derived based on the following two observations.

Observation 1: Objective function (4.2) consists of terms which are multiplication of the two assignment variables over two different periods. Hence, it is nonlinear and makes the solutions to different periods dependent on each other. However, if the initialization cost C_I and hardware cost C_A are zero, the objective function becomes linear, and the profit for each time period will not affect each other.

This provides us with a possible two-phase approach to solve the nonlinear mixed integer programming problem. In phase 1 we first set C_I and C_A both to zeros so that the problem can be separated into multiple independent subproblems and each involved a solution to a linear programming problem. In phase 2 we then add in the effect of C_I and C_A, by examining the number of new APs and APs need to be relocated.

Since the presence of C_I and C_A makes us unable to sum up the optimal solutions at different periods to obtain the overall optimal, we have to look for a larger solution space at each period rather than just the optimal solution. In order to obtain the long term optimal solution, all solutions which satisfy the constraints in each period should be used and not just the optimal solution in each period. The combinations of the feasible solutions in different periods need to be tested before the optimal can be obtained. This approach, however, will result in very large search space especially when the number of feasible solutions at each period is large (if a large number of candidate sites is used).

The total number of searches at each period of time for a given n candidate APs can be as high as $C_n^1 + C_n^2 + \cdots + C_n^n = 2^n$. Supposing the number of feasible solutions for period p is given by $l_p < 2^n$, over P periods, the combinations need to be tested is given by $\prod_{p-1}^{P} l_p$. Fortunately, the acceptable number of solutions can be much lesser by removing those feasible solutions which have unrealistically turned on too many APs, because under-utilization of the network will definitely result in lower profit. We will explain this in the following.

When performing the search, the 2^n solutions actually can form a binary tree structure, like the one shown in Figure 5-1 for a case having three binary variables. For a normal MILP model, the optimal solution is found by a branch and cut (B&C) algorithm. For example, if C is the first recorded solution satisfying the constraints, and E also satisfying all constraints but has an objective function less than C (for maximization problem), the whole branch which consists of E and K (and onwards) needs not be checked at all. However, in our case, we want to find all the possible solutions before the right combination of APs placements over different periods is obtained. The B&C algorithm needs to be modified since E, and even K, is also a possible solution. We next look into ways to limit the number of feasible solutions to be involved when searching for the final optimal solution.

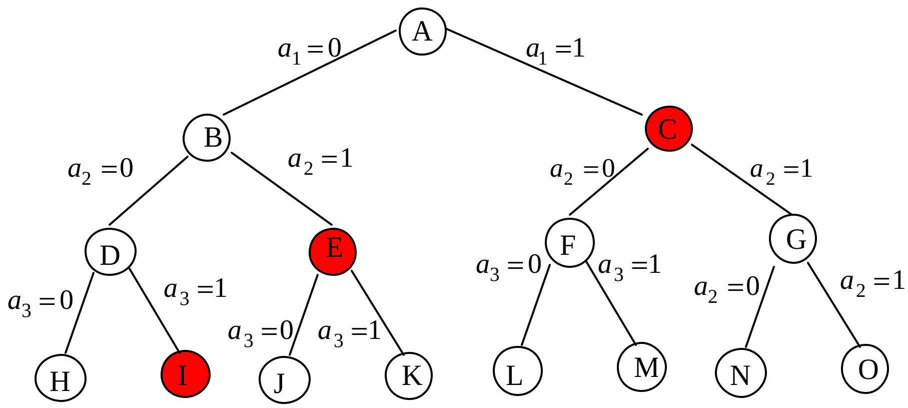


Figure 4-1 A modified branch-and-cut algorithm

Observation 2: In each period, if a solution set has a subset of APs which can result in a lower profit, then the solution set can be excluded from the checklist.

This observation is to claim that if C is a feasible solution, then G and O need not be included for the searching of optimal solution over multiple periods. The reason is G and O will definitely result in a larger cost and hence lower profit regardless of ways with which it is paired with solutions in other periods. Similarly if E is included in the search then K needs not have to. Hence, in this example, there are 7 feasible solutions (C, G, O, M, E, K, I) but only the three marked in red need to be involved in the checklist when performing optimization. The search space now becomes much smaller, especially when the number of candidate sites is high.

The proof this is straightforward by referring to objective function given in (4.2) and will not be discussed.

C. The algorithm

Based on the observations, we propose an algorithm as followed.

Step 1: For each period, obtain a checklist of solutions at each period by first eliminating C_I and C_A from the objective function.

Step 2: To obtain the projected optimal solution for the multiple periods.

Example 4.1: Let us consider a case when in period 1, the information about the DPs are given in Table 4.1, whilst in period 2, the information about the DPs is similar to the case given in Example 2.1. In both cases, the candidate AP sites and various costs are the same as that in Example 2.1.

We find the lists of solutions for period 1 (labeled as M1(.)) and period 2 (labeled as M2(.)) and tabulated them in Table 4.2.

DP	X	Y	Demand	DP	X	Y	Demand
1	3.67	4.69	0	2	5.12	1.97	0
3	3.35	4.09	0	4	1.94	1.02	1.2
5	6.00	1.84	0	6	9.44	8.03	0
7	5.50	8.71	0.1	8	4.67	2.36	0
9	6.62	4.76	0	10	4.20	1.77	0
11	8.35	5.44	0	12	3.49	0.56	0
13	5.40	3.90	0	14	4.12	5.85	1.77
15	4.89	1.85	0				

Table 4.1 Demand traffic at each demand point (T_j)

Period 1			Period 2		
Solution	R	APs selected	Solution	R	APs selected
M1(1)	-60	7 5 4	M2(1)	-89	8 7 5 4
M1(2)	-40	7 3	M2(2)	-69	8 7 3
M1(3)	-60	5 4 1	M2(3)	-89	8 5 4 1
M1(4)	-40	3 1	M2(4)	-69	8 3 1

Table 4.2 Demand traffic at each demand point (T_j): Feasible solutions in Period 1 and 2 to be used for optimal search

Since we have a checklist of solutions for each period, the optimal solution for the whole periods can be easily obtained by calculating the profit for each possible combination. There are a total of 16 combinations and the optimal is finally evaluated to be: activates AP 3 and 7 in period 1 and activates AP 8 while keeping AP 3 and 7 in period 2. The computed profit is -556.51.

CHAPTER 5

OPTIMIZATION MODEL WITH PROBABILISTIC FUTURE

In the previous chapter, we have discussed one model to deal with future demands. However, it is sometimes very hard to give a clear picture of the future situation.

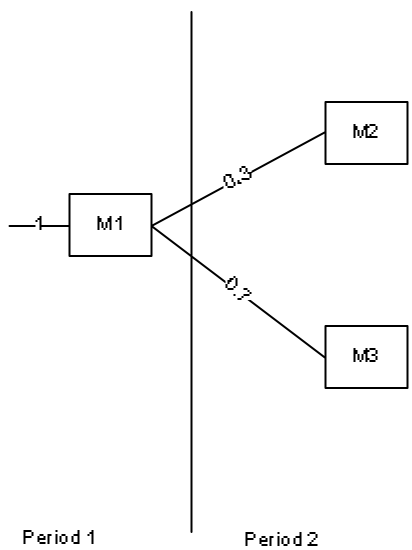
Decision analysis has become an important technique for decision making when facing uncertainty. It is performed by enumerating all the available actions, identifying the payoffs for all possible outcomes, and quantifying the subjective probabilities for all the possible random events. When these data are available, decision analysis becomes a powerful tool for determining an optimal action. Our network optimization model will use the decision analysis method to find the optimal solution over multiple time periods.

In this chapter (Chapter 5), a network cell planning model considering future predicted demand traffic is proposed. Instead of solving the deterministic demand map optimization problems at a given period of time, we illustrated a solution procedure to obtain the largest profit over multiple time periods each with different deterministic demand traffic. Without sacrificing any accuracy, a two-step optimization procedure is proposed. In the first step, a modified branch and cut algorithm to reduce the possible solutions so as to be used in each time period when evaluating the final optimal solution. The decision analysis is also used to analyze the case when probabilistic future demand traffic. Finally, the ways with which utility theory can be applied to account for individual's risk taking preference is also discussed.

5.1 Formulation with decision analysis

Decision analysis is the discipline of evaluating complex alternatives in terms of values and uncertainty. In Chapter 4, the optimal solution over a few periods with known projected traffic demands can be obtained. However, sometimes the future traffic demands cannot be predicted with 100% surety. Thus, we need to develop new decision process to compute the solution which generates the "best expected" profit by modifying the algorithm developed in Chapter 4.

Figure 5-1 gives an example with two periods. M1 is the known demand traffic in current period. Period 2 has two possible projected traffic demands due to uncertainty in whether the services will take off. The probabilities of their occurrence are given besides the links as 0.3 and 0.7. In this section, because we now have different traffic demand in M2 and M3, hence when making decision, the revenue term cannot be removed like in previous two sections. In our example, we take P =30 be the charge per unit traffic. We will add in the revenue in step 2 of the modified algorithm.



DP	X	Y	Demand	DP	X	Y	Demand
1	3.67	4.69	0	2	5.12	1.97	3.0
3	3.35	4.09	2	4	1.94	1.02	0
5	6.00	1.84	0	6	9.44	8.03	0
7	5.50	8.71	0	8	4.67	2.36	2.86
9	6.62	4.76	0	10	4.20	1.77	0
11	8.35	5.44	0	12	3.49	0.56	0
13	5.40	3.90	3	14	4.12	5.85	0
15	4.89	1.85	1				

Table 5.1 Traffic demand M3 in the 2nd period

Solution	*R*	APs selected
M3(1)	-40	8 7
M3(2)	-20	4
M3(3)	-40	7 2
M3(4)	-40	8 1
M3(5)	-40	2 1

Table 5.2 Solutions for M3 in the 2nd period

For each of the solutions in M1, we need to find out one solution in the following period which gives the best profit R over the two periods. These results are tabulated in Table 5.3. Using the two solutions (M1 to M2 and M1 to M3 given), the possible average profits for the two periods are next tabulated in the last column of Table 5.3.

Solution in M1	Optimal solution in M2	Optimal profit	Solution in M1	Optimal solution in M3	Optimal profit	Expected Profit
M1(1)	M2(1)	538.57	M1(1)	M3(2)	-94.36	95.519
M1(2)	M2(2)	729.59	M1(2)	M3(1)	-59.61	177.15
M1(3)	M2(3)	489.48	M1(3)	M3(2)	-163.45	32.429
M1(4)	M2(2)	720.50	M1(4)	M3(4)	-68.70	168.06

Table 5.3 Possible optimal solutions given a solution of M1 (a) M2 (b) M3 (c) weighted sum

Figure 5-2 summarizes the results shown in Table 5.3. From the above solution process, we can find that solution 2 in the first period has the best expected profit equal to 177.15. The problem takes 26 minutes to solve using an Intel 1.5GHz Linux PC.

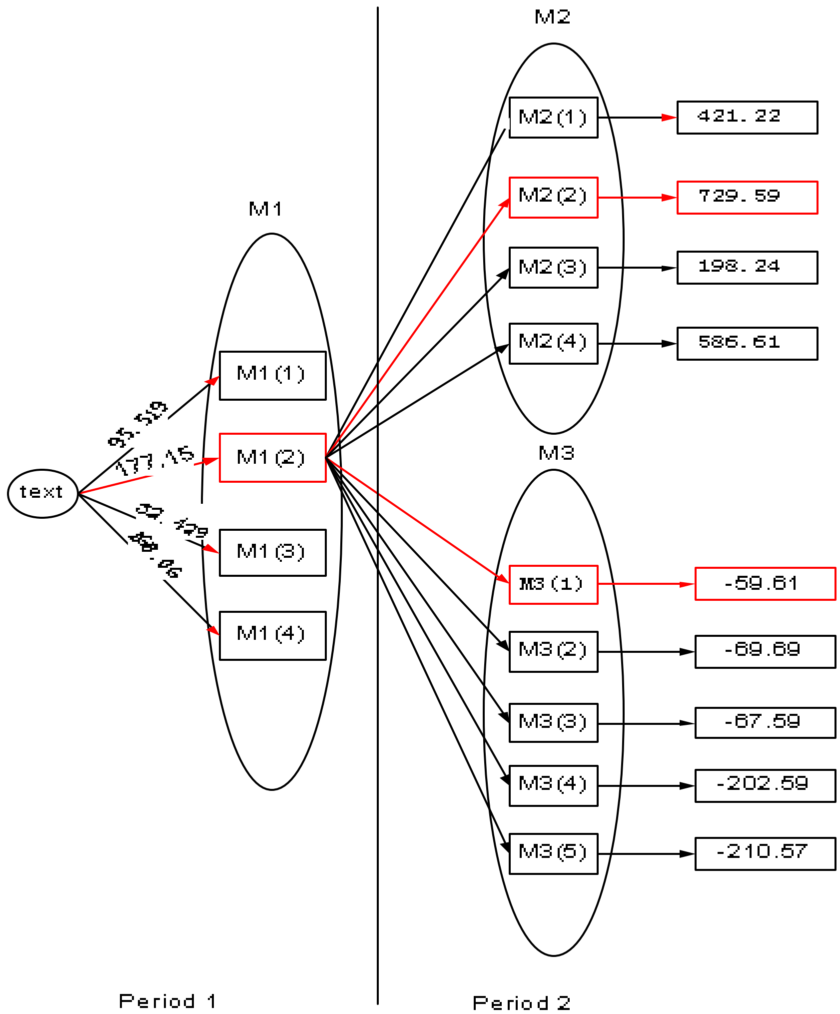


Figure 5-2 Evaluation of optimal solution for probabilistic traffic demand

5.2 Use of utility theory in decision making

In the above decision analysis, we assumed that the expected revenue expressed in monetary terms is the appropriate measure of the consequences of taking an action. However, in many situations this assumption is inappropriate. For example, SPs may want to take a risk to get a possible higher return.

The problem is then a need to have risk taking measurement. The utility theory is used to serve this purpose. It is a way of transforming monetary values to an appropriate scale that reflects the decision maker's preferences. In utility theory, a utility has to be assigned to each of the possible (and mutually exclusive) consequences of every alternative. A utility function is the rule by which this assignment is done and its choice depends on the preferences of the individual decision maker.

Using the previous example, assume that the SPs have the exponential utility function:

$$u(R) = r(1 - e^{-\frac{R}{r}}) \tag{5.1}$$

where $u(R)$ indicates the corresponding utility of having profit R. r is the decision maker's risk tolerance. This utility function has a decreasing marginal utility for money, so it is a designed to fit a risk-averse individual. A great aversion to risk corresponds to a small value of r (which would cause the utility function curve to bend sharply). Decision maker needs to find the suitable r for himself. For r = 800, the utility function is shown in Figure 5-3. If the decision maker now makes decision based on his utility function, he can use the utilities derived from each of the profit to make his decision, rather than using R. The results after transforming from Table 5.2 are shown in Table 5.4.

From the calculated result, the utility of selecting solution 2 for M1 still gives the largest predicted profit. Hence, solution 2 is still the best option. With the utility theory model, the SP can have a better decision which is suitable to his own marketing strategy.

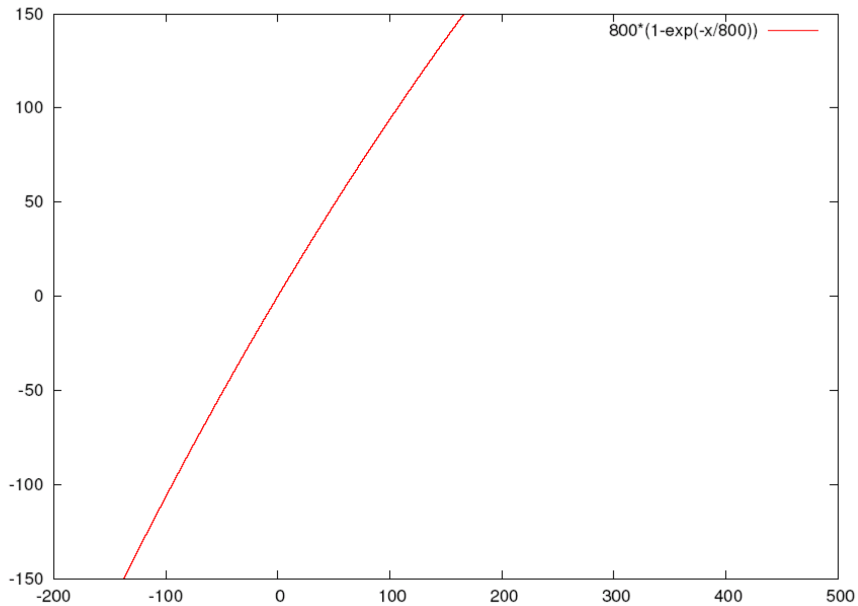


Figure 5-3 An utility function: $u(R)$ versus R

Solution	Utility (M1-M2,30%)	Utility (M1-M3, 70%)	Expected utility
M1(1)	391.94	-100.15	47.4768
M1(2)	478.62	-61.887	100.2651
M1(3)	366.12	-181.345	-17.1052
M1(4)	474.94	-71.7361	92.2689

Table 5.4 Utility values for various solution

CHAPTER 6

CONCLUSIONS AND FUTURE RESEARCH

6.1 Concluding remarks

Wireless network optimization has long been researched and discussed over the past two decades. In this thesis, a platform to look into the deployment of wireless networks which is able to optimize the profit generated over multiple periods, each with different spatial traffic demand, is developed.

In Chapter 2 and Chapter 3 of this thesis, we model the optimal cell placement and coverage problem with mixed linear integer programming model, both for the FDMA based and CSMA/CA based networks. Given a set of candidate sites, we first derive the placement and compute the transmission power of the access points to support a given spatial traffic demand over a specific period of time. Adjustable transmission range is made possible through power control to minimize the amount of interference among neighbouring access points.

From practical viewpoints, there are more factors still need to be taken into consideration. For example, the deployment of wireless networks needs to consider longer term investment return. When a new wireless network is just to be deployed, the service providers would like to do it in a progressive and conservative way. The reason is because the technology will take some time to take off. The service providers will be interested to find a way to minimize the cost or maximize profit. To our best knowledge, no research has been made in this aspect. The gradual increase of demand traffic imposes a challenge to the design: closes and reopening of APs so as to match the changes in traffic demand over different periods of time. The solution for a long term profit is a challenging problem because optimized solution in each period may not necessary guarantee overall optimal. With the knowledge on the projected demand traffic in subsequent periods, algorithms to maximize the long term profit are developed when the projected traffic are deterministic in Chapter 4. Numerical results showed that the overall search space for solution can be quite large which makes it impossible to solve in a reasonable time. Thus we developed new branch-and-cut algorithm to simplify the solution process. The main idea of the algorithm is to eliminate the obvious infeasible solution and hence the time of solving such problems are shortened to a great extent.

Finally, the prediction of future traffic cannot be accurately made as a result of competition from other SPs and the acceptability of services. We feel that a more proper way of planning is to look for possible future traffic demands each with a given probability of occurrence. In Chapter 5, the design challenges to extend the algorithm to maximize profit over multiple periods with probabilistic demand traffic are given. The use of utility theory in such scenario is also discussed to adapt to different user needs.

6.2 Future research

This thesis illustrates the way to use operational research in wireless network optimization. However, there still exists unsatisfactory in the solution process. These include:

1) Calculation time. Although many commercial and free linear programming solvers are available in the market, when the number of variables increases, the time to calculate such linear or integer programming problems grows exponentially. This is the normally encountered bottleneck to obtain the solutions. Many researches have been carried to develop algorithms which can speed up the solution process. However, there seems to be much space to be explored.
2) The use of saturation throughput is just an approximate approach. It remains a challenge to include the CSMA/CA protocol in the OR model, which include other factors which affect the throughput, such as the packet size, number of mobile stations and the backoff algorithm.

PUBLICATION

[1] Q. M. Wu, Y. H. Chew and B. S. Yeo, " Multi-Periods Optimization Strategy for Wireless Network Deployment", *IEEE Wireless Communications and Networking Conference*, Mar, 2007

REFERENCES

[1] W. R. Young, "AMPS: Introduction, Background, and Objectives", *Bell System Technical Journal*, vol. 58, 1, pp. 1-14, Jan 1979.

[2] S. M. Redl, M. K. Weber and W.M. Oliphant , "An Introduction to GSM", *ISBN 978-0890067857*, Artech House, Mar 1995

[3] M. Sauter, “Communication Systems for the Mobile Information Society”, *John Wiley*, *ISBN 0-470-02676-6*, Sep 2006,

[4] IEEE 802.11 Working Group (2007-06-12). “IEEE 802.11-2007: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications”, *ISBN 0 -7381-5656-9*.

[5] T. S. Rappaport, “Wireless Communications: Principles and Practice”, *Second Edition, Prentice Hall*,2002

[6] S. Ceria, P. Nobili and A. Sassano, “Set covering problem”, *Annotated Bibliographies in Combinatorial Optimization*, John Wiley, Ch. 23, 1997.

[7] A. Hills, “Large-scale wireless LAN design”, *IEEE Communications Magazine*, vol. 39, no. 11, pp. 98-107, Nov 2001.

[8] E. Amaldi, A. Capone, M. Cesana, F. Malucelli and F. Palazzo, “Optimizing WLAN radio coverage”, *IEEE International Conference on Commun*., pp.2219-2223, Jun. 2004

[9] S. Hurley, “Planning Effective Cellular Mobile Radio Networks”, *IEEE Trans. on Vech. Technol.*, vol. 51, no 2, pp.243-253, Mar. 2002.

[10] R. C. Rodrigues, G. R. Mateus and A. A. F. Loureiro, "On the design and capacity planning of a wireless local area network", in Proc. *IEEE/IFIP Network Operations and Management Symp*. (NOMS 2000).

[11] Y. Lee, K. Kim and Y. Choi, “Optimization of AP placemment and channel assignment in wireless LANs”, *IEEE conference on Local Computer Networks* (*LCS 2002*), pp. 6, 2002.

[12] Y. Ngadiman, Y. H. Chew and B. S. Yeo “A new approach for finding optimal base stations configuration for CDMA systems jointly with uplink and downlink constraints”, *IEEE Personal, Indoor and Mobile Radio Communications* (*PIMRC*), 2005.

[13] H. H. Liu, J. Lien and C. Wu, “A scheme for supporting voice over IEEE 802.11 wireless local area network”, *Proc. National Science Council. ROC*(*A*), vol. 25, no. 4, 2001. pp. 259-268.

[14] Z. H. Velkov and B. Spasenovski, “An analysis of CSMA/CA protocol with capture in wireless LANs”, *IEEE Wireless Communications and Networking* (*WCNC*), 2003.

[15] J. H. Kim and J. K. Lee, “Capture effects of wireless CSMA/CA protocols in Rayleigh and shadow fading channels”, *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, Jul 1999.

[16] J. Jun, P. Peddabachagari and M. L. Sichitiu, “Theoretical maximum throughput of IEEE 802.11 and its applications”, *IEEE International Symposium on Network Computing and Applications* (*NCA*), 16-18 Apr 2003.

[17] W. Yue and Y. Matsumoto, “An exact analysis for CSMA/CA protocol in integrated voice / data wireless LANs”, *IEEE Global Telecommunications Conference* (*GLOBECOM*), 2000.

[18] T. Fujita, T. Onizawa, S. Hori, A. Ohta and S. Aikawa,. “An evaluation scheme of cell throughput for multi-rate wireless LAN systems with CSMA/CA”, *IEEE 58th Vehicular Technology Conference* (*VTC Fall*), Fall 2003.

[19] S. Choudhury and J. D. Gibson, "Payload length and rate adaptation for throughput optimization in wireless LANs", *IEEE 63rd Vehicular Technology Conference* (*VTC*), Spring 2006.

[20] G. Bianchi, "Performance analysis of the IEEE 802.11 Distributed Coordination Function" , *IEEE Journal Selected Areas in Communications*, vol 18, no 3, Mar 2000.

[21] X. Wang and K. Kar, "Throughput modeling and fairness issues in CSMA/CA based Ad-Hoc networks" , *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, (*INFOCOM*), vol 1, pp.23-34, Mar 2005

[22] J. K. Anderson and N. Youell, "A closer look at WLAN throughput and performance", *Bechtel Telecommunications Technology Group* http://www.bechteltelecoms.com/docs/bttj_v1/Article14.pdf

[23] R. L. Abrahams, "Evaluating wideband 802.11 WLAN radio performance", *Harris Corp COMMSDESIGN*, Mar 2004.

[24] Link Planning tools for Wireless LAN (WLAN), available in (http://huizen.deds.nl/~pa0hoo/helix_wifi/linkbudgetcalc/wlan_budgetcalc.html)