Intelligent Radio Resource Management for WLANs

Yapeng Wang

Submitted for the degree of Doctor of Philosophy

Department of Electronic Engineering
Queen Mary, University of London

2007
To my beloved wife and my parents
Abstract

The demand for IEEE 802.11 based WLANs has grown quickly over the recent few years and while new standards (e.g. 802.11 a/g) are being developed, the wireless link is still the bottleneck in WLAN systems, especially in a public WLAN deployment with high-demand users. The most common way of improving the system performance is by accurate site planning, but this still does not cope with dynamic demand from users, such as in hotspots at an airport where travellers with WLAN laptops tend to gather at different gates at different times. In this thesis, a novel intelligent radio resource management system for WLANs with low-cost semi-smart antennas is proposed and evaluated. The access points can re-configure their radio coverage patterns cooperatively to balance hot-spot traffic. This work is the first to consider such technology in an indoor environment.

In this thesis, the concept of using semi-smart antennas for large scale WLAN systems is introduced and a full-scale system-level WLAN simulator is developed to simulate large scale WLAN deployment with different user scenarios. Using an agent-based negotiation system to dynamically change the radio coverage patterns, the simulation results show significant performance improvement over traditional configurations. Since most WLANs are deployed indoors, the multi-path propagation effects must be considered and the initial negotiation algorithm is enhanced to cope with this indoor, multi-path environment.

Finally, the distributed agent negotiation approach is compared with a global optimisation algorithm that allows the overhead of the control algorithms to be reduced.

All results have demonstrated that significant capacity improvement can be gained by the use of this intelligent radio resource management system for WLANs.
Acknowledgement

I would like to express my deep and sincere gratitude to my supervisor, Professor Laurie Cuthbert. His wide knowledge and his logical way of thinking have been of great value for me. His understanding, encouraging and personal guidance have provided the basis for the present thesis.

I would like to thank Prof Clive Parini, Dr John Bigham, Dr Athen Ma, Dr Lin Du (graduated PhD student), Dr Yue Chen, Dr Na Yao and many others in Queen Mary who have given me kind help and valuable suggestions and comments during my PhD. I would also like to thank people from administrative and technical support department - Melissa Yeo, Lynda Rolfe, Michele Pringle, Kok Ho Huen and many others for their help during my studies.

I would like to thank all my friends at Queen Mary for the good times and friendship.

Finally, I would like to thank my dearest wife, Xu Yang, who gave me enormous help and encouragement all the time, and my parents and parents-in-law for their love and support.
Contents

ABSTRACT ............................................................................................................................................... 3
ACKNOWLEDGEMENT ....................................................................................................................... 4
CONTENTS ............................................................................................................................................... 5
LIST OF FIGURES .................................................................................................................................... 7
ABBREVIATIONS ..................................................................................................................................... 9

CHAPTER 1 INTRODUCTION .................................................................................................................. 11
  1.1 INTRODUCTION ........................................................................................................................... 11
  1.2 SCOPE OF THE RESEARCH ......................................................................................................... 12
  1.3 RESEARCH CONTRIBUTIONS ...................................................................................................... 13
  1.4 AUTHOR’S PUBLICATIONS ......................................................................................................... 13
  1.5 ORGANISATION OF THIS THESIS ............................................................................................ 14

CHAPTER 2 BACKGROUND ..................................................................................................................... 16
  2.1 INTRODUCTION ........................................................................................................................... 16
  2.2 THE IEEE 802.11 STANDARD ..................................................................................................... 17
  2.2.1 802.11 MAC ............................................................................................................................ 21
  2.2.2 IEEE 802.11 Physical layer .................................................................................................... 26
  2.3 RRM FOR WIRELESS NETWORKS .............................................................................................. 30
  2.3.1 Smart antenna research ........................................................................................................... 31
  2.4 EXTENDING RADIO RESOURCE MANAGEMENT (RRM) FUNCTIONS TO 802.11 NETWORKS........... 34
  2.4.1 Introduction ............................................................................................................................ 34
  2.4.2 The common WLAN deployment practice ............................................................................. 34
  2.4.3 Limitations of the common practice and various RRM schemes .......................................... 37
  2.5 SUMMARY .................................................................................................................................... 38

CHAPTER 3 NEW RRM FRAMEWORK FOR 802.11 WLANS ................................................................. 40
  3.1 A SEMI-SMART SECTORED ANTENNA FOR 802.11 AP ......................................................... 40
  3.2 NEW ELEMENT ARRAY ANTENNA FOR 802.11 WLAN ........................................................... 42
  3.2.1 Ekahau position engine for locating WLAN users ................................................................... 48
  3.3 MONITORING TRAFFIC LOADS OF WLAN ACCESS POINTS ..................................................... 49
  3.3.1 Introduction ............................................................................................................................ 49
  3.3.2 Investigating MIBs for 802.11 Access Point and Client ......................................................... 51
  3.4 SUMMARY .................................................................................................................................... 54

CHAPTER 4 INTELLIGENT AGENT APPROACH TO RRM ................................................................. 55
  4.1 INTELLIGENT AGENTS ............................................................................................................... 55
  4.2 MULTI-AGENT SYSTEMS ........................................................................................................... 57
  4.3 AGENT PLATFORM ....................................................................................................................... 58
  4.4 AGENT-BASED NEGOTIATION FRAMEWORK ........................................................................... 62
  4.4.1 Architecture ........................................................................................................................... 62
  4.4.2 Example negotiation ............................................................................................................... 64
  4.4.3 Negotiation process ............................................................................................................... 67
  4.4.4 Commitment process ............................................................................................................. 71
  4.5 SIMULATOR FOR WLAN SYSTEM WITH SEMI-SMART ANTENNAS ......................................... 73
  4.5.1 Introduction ........................................................................................................................... 73
  4.5.2 System architecture ................................................................................................................. 74
  4.5.3 Simulation validation .............................................................................................................. 77
  4.5.4 Measurement and pass loss calculation hybrid method .......................................................... 80
  4.6 THE SIMULATION FOR THE AGENT-BASED APPROACH ....................................................... 81
  4.6.1 The Result ............................................................................................................................. 82
  4.7 SUMMARY .................................................................................................................................... 87

CHAPTER 5 SIMULATION UNDER CONSTRAINED ENVIRONMENT ................................................... 88
  5.1 INTRODUCTION ........................................................................................................................ 88
  5.2 DATA RATES ACHIEVED IN A MULTI-USER ENVIRONMENT ................................................... 89
  5.3 MODELLING WLAN USERS TRAFFIC ...................................................................................... 91
5.4 MODIFICATION AND ALTERNATION TO THE NEGOTIATION ALGORITHM ........................................ 94
5.5 MULTI-PATH RADIO PROPAGATION .................................................................................................. 96
  5.5.1 Using site-survey to generalise the radio propagation map ......................................................... 97
  5.5.2 Using software simulation to produce radio propagation map .................................................... 98
5.6 APPLYING RADIATION PATTERN FROM A SEMI-SMART ANTENNA ................................................ 98
5.7 SIMULATING A PSEUDO REAL SCENARIO .................................................................................... 100
5.8 OBSTACLES IN AN INDOOR ENVIRONMENT ................................................................................ 107
5.9 SUMMARY ...................................................................................................................................... 109

CHAPTER 6 GLOBAL OPTIMISATION ALGORITHM .............................................................................. 111
  6.1 INTRODUCTION ........................................................................................................................... 111
  6.2 GLOBAL OPTIMISATION FOR CONSTRAINED 802.11 WLANS .................................................... 111
  6.3 SIMULATION RESULTS ................................................................................................................ 119
    6.3.1 Comparison with negotiation ................................................................................................... 119
  6.4 DIFFERENT HOTSPOT FORMATION ............................................................................................. 122
  6.5 COMPLEX MULTI-PATH CONSTRAINED ENVIRONMENT ............................................................ 124
  6.6 SUMMARY .................................................................................................................................... 132

CHAPTER 7 CONCLUSION AND FUTURE WORK ................................................................................. 134
  7.1 CONCLUSION ............................................................................................................................... 134
  7.2 FURTHER WORK ......................................................................................................................... 134

AUTHOR’S ADDITIONAL PUBLICATIONS ..................................................................................... 136
REFERENCES ....................................................................................................................................... 137
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Ad Hoc and Infrastructure modes</td>
<td>18</td>
</tr>
<tr>
<td>2.2</td>
<td>BSS transition</td>
<td>19</td>
</tr>
<tr>
<td>2.3</td>
<td>Reassociation process in detail</td>
<td>20</td>
</tr>
<tr>
<td>2.4</td>
<td>Nodes 1 and 3 are 'hidden'</td>
<td>22</td>
</tr>
<tr>
<td>2.5</td>
<td>Using NAV for virtual carrier-sensing</td>
<td>23</td>
</tr>
<tr>
<td>2.6</td>
<td>An example timing of 802.11 DCF</td>
<td>25</td>
</tr>
<tr>
<td>2.7</td>
<td>Interframe spacing relationships</td>
<td>25</td>
</tr>
<tr>
<td>2.8</td>
<td>Overlapping channels of 802.11</td>
<td>28</td>
</tr>
<tr>
<td>2.9</td>
<td>Expected 802.11a, 802.11b and 802.11 g data rates at varying distance from access point [42]</td>
<td>30</td>
</tr>
<tr>
<td>2.10</td>
<td>Illustration of smart antenna types</td>
<td>32</td>
</tr>
<tr>
<td>2.11</td>
<td>Prototype semi-smart antenna and its beamforming</td>
<td>33</td>
</tr>
<tr>
<td>2.12</td>
<td>Manual WLAN deployment process</td>
<td>35</td>
</tr>
<tr>
<td>2.13</td>
<td>Frequency reuse pattern for 801.11b/g</td>
<td>36</td>
</tr>
<tr>
<td>3.1</td>
<td>The appearance and working diagram of the four-sector ceiling mounted antenna array</td>
<td>42</td>
</tr>
<tr>
<td>3.2</td>
<td>Coverage patterns of the antenna sectors with different power control</td>
<td>43</td>
</tr>
<tr>
<td>3.3</td>
<td>Cooperative changing AP coverage patterns to reduce traffic hot spot</td>
<td>44</td>
</tr>
<tr>
<td>3.4</td>
<td>illustrates the angles required for coverage of users</td>
<td>44</td>
</tr>
<tr>
<td>3.5</td>
<td>a) the fabricated antenna; b) mounted on the four faces of the pyramid; and c) the theoretical (simulation) and measured radiations pattern</td>
<td>45</td>
</tr>
<tr>
<td>3.6</td>
<td>triangulation to locate a mobile station</td>
<td>46</td>
</tr>
<tr>
<td>3.7</td>
<td>Time difference of signal pulses from AP to STA (Fig on page 6 of [76])</td>
<td>47</td>
</tr>
<tr>
<td>3.8</td>
<td>Tracking mode of Ekahau</td>
<td>49</td>
</tr>
<tr>
<td>3.9</td>
<td>Structure for managing WLAN network</td>
<td>51</td>
</tr>
<tr>
<td>3.10</td>
<td>MIB-II [82]</td>
<td>53</td>
</tr>
<tr>
<td>4.1</td>
<td>Structure of utility-based agent</td>
<td>57</td>
</tr>
<tr>
<td>4.2</td>
<td>FIPA agent management reference model</td>
<td>59</td>
</tr>
<tr>
<td>4.3</td>
<td>FIPA compliant Communication model</td>
<td>60</td>
</tr>
<tr>
<td>4.4</td>
<td>Software architecture of the JADE Platform</td>
<td>62</td>
</tr>
<tr>
<td>4.5</td>
<td>The Multiple Agent System for WLAN resource management</td>
<td>63</td>
</tr>
<tr>
<td>4.6</td>
<td>Access Point Agent and Antenna Agent basic structure</td>
<td>63</td>
</tr>
<tr>
<td>4.7</td>
<td>The basic negotiation sequence diagram</td>
<td>66</td>
</tr>
<tr>
<td>4.8</td>
<td>Polar coordinates represent AP’s coverage and two example hypotheses</td>
<td>68</td>
</tr>
<tr>
<td>4.9</td>
<td>State transitions of coverage negotiation</td>
<td>70</td>
</tr>
<tr>
<td>4.10</td>
<td>State transitions of commitment</td>
<td>72</td>
</tr>
<tr>
<td>4.11</td>
<td>flow chart of the WLAN network simulator</td>
<td>76</td>
</tr>
<tr>
<td>4.12</td>
<td>AP Layout of the simulation</td>
<td>82</td>
</tr>
<tr>
<td>4.13</td>
<td>Simulation result of using static user demand</td>
<td>83</td>
</tr>
<tr>
<td>4.14</td>
<td>Simulation result of using variable user demand</td>
<td>84</td>
</tr>
<tr>
<td>4.15</td>
<td>Screen shot of part of the simulation GUI</td>
<td>84</td>
</tr>
<tr>
<td>4.16</td>
<td>Simulation result of and comparison of 802.11b and 802.11g (variable user demand)</td>
<td>85</td>
</tr>
<tr>
<td>4.17</td>
<td>Improvement obtained with negotiation (802.11b)</td>
<td>86</td>
</tr>
<tr>
<td>4.18</td>
<td>Improvement obtained with negotiation (802.11g)</td>
<td>87</td>
</tr>
<tr>
<td>5.1</td>
<td>the exponential model of 802.11 b data rate approximation</td>
<td>90</td>
</tr>
<tr>
<td>5.2</td>
<td>Traffic model for user demand</td>
<td>92</td>
</tr>
<tr>
<td>5.3</td>
<td>Download speeds of various HTTP and FTP traffic session</td>
<td>93</td>
</tr>
<tr>
<td>5.4</td>
<td>QC based negotiation</td>
<td>95</td>
</tr>
</tbody>
</table>
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2G</td>
<td>Second Generation (Mobile)</td>
</tr>
<tr>
<td>3G</td>
<td>Third Generation (Mobile)</td>
</tr>
<tr>
<td>3GPP</td>
<td>3G Partnership Project</td>
</tr>
<tr>
<td>ACC</td>
<td>Agent Communication Channel</td>
</tr>
<tr>
<td>ACL</td>
<td>Agent Communication Language</td>
</tr>
<tr>
<td>AFH</td>
<td>Ask For Help</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>AMS</td>
<td>Agent Management System</td>
</tr>
<tr>
<td>AOC</td>
<td>Acknowledgement of Commitment</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>BSS</td>
<td>Basic Service Set</td>
</tr>
<tr>
<td>CCK</td>
<td>Complementary Code Keying</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code Division Multiple Access</td>
</tr>
<tr>
<td>CFP</td>
<td>Contention Free Period</td>
</tr>
<tr>
<td>CMP</td>
<td>Change My Pattern</td>
</tr>
<tr>
<td>CORBA</td>
<td>Common Object Request Broker Architecture</td>
</tr>
<tr>
<td>CP</td>
<td>Contention Period</td>
</tr>
<tr>
<td>CSMA/CA</td>
<td>Carrier Sense Multiple Access with Collision Avoidance</td>
</tr>
<tr>
<td>CSMA/CD</td>
<td>Carrier Sense Multiple Access with Collision Detection</td>
</tr>
<tr>
<td>CT</td>
<td>Cancel Transaction</td>
</tr>
<tr>
<td>CTS</td>
<td>Clear to Send</td>
</tr>
<tr>
<td>CW</td>
<td>Contention Window</td>
</tr>
<tr>
<td>CYP</td>
<td>Change Your Pattern</td>
</tr>
<tr>
<td>DBPSK</td>
<td>Differential Binary Phase Shift Keying</td>
</tr>
<tr>
<td>DCF</td>
<td>Distributed Coordination Function</td>
</tr>
<tr>
<td>DF</td>
<td>Directory Facilitator</td>
</tr>
<tr>
<td>DIFS</td>
<td>Distributed Inter-Frame Space</td>
</tr>
<tr>
<td>DQPSK</td>
<td>Differential Quadrature Phase Shift Keying</td>
</tr>
<tr>
<td>DS</td>
<td>Direct Sequence or Distributed System</td>
</tr>
<tr>
<td>EDCA</td>
<td>Enhanced DCF Channel Access</td>
</tr>
<tr>
<td>ESS</td>
<td>Extended Service Set</td>
</tr>
<tr>
<td>FH</td>
<td>Frequency Hopping</td>
</tr>
<tr>
<td>FIPA</td>
<td>Foundation for Intelligent Physical Agents</td>
</tr>
<tr>
<td>GOA</td>
<td>Global Optimisation Algorithm</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communication</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>HCCA</td>
<td>HCF Controlled Channel Access</td>
</tr>
<tr>
<td>HCF</td>
<td>Hybrid Coordination Function</td>
</tr>
<tr>
<td>HIPERLAN</td>
<td>High Performance Radio LAN</td>
</tr>
<tr>
<td>IAPP</td>
<td>Inter-Access Point Protocol</td>
</tr>
<tr>
<td>IIOP</td>
<td>Internet Inter-Orb Protocol</td>
</tr>
<tr>
<td>IR</td>
<td>Infra-red</td>
</tr>
<tr>
<td>ISM</td>
<td>Industrial Scientific and Medical</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MAS</td>
<td>Multi-Agent System</td>
</tr>
<tr>
<td>MIB</td>
<td>Management Information Base</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple-Input Multiple-Output</td>
</tr>
<tr>
<td>MS</td>
<td>Mobile Station</td>
</tr>
<tr>
<td>NAV</td>
<td>Network Allocation Vector</td>
</tr>
<tr>
<td>NMS</td>
<td>Network Management Station</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency-Division Multiplexing</td>
</tr>
<tr>
<td>OMG</td>
<td>Object Management Group</td>
</tr>
<tr>
<td>PCF</td>
<td>Point Coordination Function</td>
</tr>
<tr>
<td>PHY</td>
<td>Physical (layer)</td>
</tr>
<tr>
<td>PIFS</td>
<td>PCF interframe space</td>
</tr>
<tr>
<td>PWLAN</td>
<td>Public Wireless LAN</td>
</tr>
<tr>
<td>QC</td>
<td>Quantisation Cell</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RAFH</td>
<td>Reply AFH</td>
</tr>
<tr>
<td>RFC</td>
<td>Request For Commitment</td>
</tr>
<tr>
<td>RRM</td>
<td>Radio Resource Management</td>
</tr>
<tr>
<td>RSSI</td>
<td>Received Signal Strength Indication</td>
</tr>
<tr>
<td>RTS</td>
<td>Request to Send</td>
</tr>
<tr>
<td>SE</td>
<td>Scheme Expires</td>
</tr>
<tr>
<td>SNMP</td>
<td>Simple Network Management Protocol</td>
</tr>
<tr>
<td>SSL</td>
<td>Secure Socket Layer</td>
</tr>
<tr>
<td>STA</td>
<td>Station (in IEEE 802.11)</td>
</tr>
<tr>
<td>TC</td>
<td>Transaction Cancelled</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunication System</td>
</tr>
<tr>
<td>WirelessMAN</td>
<td>Wireless Metropolitan Area Networks</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
</tbody>
</table>
Chapter 1 Introduction

1.1 Introduction

802.11 based WLAN standards are a set of very successful wireless standards and WLAN products based on those standards have spread to all mobile computing devices including mobile phones. Since the initial publication in 1997 [1], a lot of research work has been undertaken, particularly for improved modulation techniques in the physical layer [2, 3] and in better efficiency in the MAC layer [2, 4].

The 802.11 standard itself has evolved to adapt faster data rates from the original 2 Mbits/s to the currently popular 802.11g that supports 54 Mbits/s; beyond that, the new standard under development (802.11n) promises even speeds up to 540 Mbits/s. Along with the development of 802.11 standards, the demand by users for high data rates to support new applications (such as Video on demand, IP TV and peer to peer downloads) has been growing rapidly.

The 802.11 was initially designed for deployment with a single access point: it tries to serve clients as long as they can associate to the access point with a minimum link quality. In the case of multiple APs forming a large overlapping coverage, mobile clients just associate to the AP that offers the strongest signal power. This works well when users are distributed fairly uniformly across the network and each AP has a moderate traffic load [5]. However, when users are bunched together to form a hot spot, and when their traffic demand is higher than the capacity of the AP serving the hot spot, the AP will be heavily loaded and users will experience lower offered data rates.

Until now, there is no standard available to solve that problem, although a few studies [6-12] have been carried out using different techniques to balance users between cells when non-uniform user and traffic distributions occur. However, these proprietary solutions mainly use association control (similar to Connection Admission Control) and need modification to the current 802.11 standards. As the 802.11 standards are already published, their solutions offer little practical benefit. Moreover, they are semi-static: association control only occurs when the user first tries to associate.

In cellular networks, however, the concept of dynamic load control is well known with
users being handed-over to other cells in order to optimise the overall network capacity. In 2G networks this is fairly straightforward as techniques such as channel borrowing [13, 14] can be used to provide capacity where it is needed, but in 3G CDMA-based networks that use a frequency reuse of factor of 1 (i.e. all cells use the same frequency allocation) different approaches are called for. One such approach is to use semi-smart antennas to control the radiation pattern in a collaborative way to match the geographical distribution of load [15].

This research investigates whether these concepts from mobile networks can be used in the WLAN environment.

1.2 Scope of the research

The scope of this research is to investigate control methods and algorithms to allow semi-smart antennas to increase the utilisation of WLAN networks and to offer mobile clients as much bandwidth as possible. This is approached using the semi-smart concept that has been applied to 3G W-CDMA networks [15-20] but for the very different environment of WLANs.

Although there is some research work in the literature [21, 22] trying to utilise fully adaptive antennas to increase 802.11 capacity, the aim there is mainly concentrating on improving air channel quality for individual users, the cooperation between APs being seldom considered. Moreover, fully adaptive antennas are considerably more expensive than the semi-smart antenna approach, which is why it is unlikely that such antennas would ever be used in WLANs where the cost of equipment is very low.

The other important aspect about WLAN deployment is that the radio environment is very different from that for cellular networks. While most researchers in mobile networks consider ideal layouts with a mesh of hexagonal cells, this would be a lot less realistic for WLANs as most installations are indoors. In such conditions, the radio propagation is very dependent on obstacles and reflections and this too is considered in this research.

The application context for the dynamic resource management came from the EU project “ADAMANT”, but the work on dynamic resource management using semi-smart antennas for WLANs is entirely that of the author.
1.3 Research contributions

The work reported in this thesis is novel. The main contributions are:

- Collaborative smart antennas for WLANs that show improved network utilisation and better throughput for individual wireless stations.
- A multi-agent based cooperative control method: a distributed control mechanism whereby each access point is controlled by a software agent (AP agent). All AP agents co-operate by exchanging messages and mutually adjusting coverage according to traffic conditions.
- A global control algorithm: a centralised control algorithm in which all APs are controlled by a central server to maximise network utilisation.
- Application of these methods to indoor environments with obstacles and severe multi-path effects.

In both control approaches, the modelling and simulation takes account of indoor propagation conditions, unlike the scenarios usually used in cellular network research. This is different from most previous research on WLAN network performance which uses simple path loss models to simulate the radio propagation, resulting in circular coverage shapes for access points (APs). The indoor environment is much more complicated, since walls and obstacles tend to absorb, reflect, and diffract signals resulting in very different coverage patterns.

To support this investigation, a simulator for evaluating the semi-smart control methods was developed. This simulator is capable of simulating multiple WLAN networks and clients at the system/application level plus the ability to adapt the semi-smart antennas. There are some commercial network simulators that can simulate 802.11 networks (e.g. OPNET [23] and J-SIM [24]) but they are generally simulating 802.11 at packet/frame level, which makes the simulation time too long for representing large scale networks. Moreover, they lack the support for simulating the proprietary semi-smart antennas used in this research.

1.4 Author's Publications

Note: these publications are directly related to the research described in this thesis; additional publications by the author are shown on page 136. These additional
publications relate to the application environment in which the dynamic resource management would be useful, rather than to the technique itself.

[Wang-1] Y. Wang, L. Cuthbert, J. Bigham, Agent-based Load Balancing of WLAN in In-door Usage, the 3rd IASTED International Conference on Wireless and Optical Communications (WOC2003), July, Banff, Canada


1.5 Organisation of this thesis

The thesis is organised as follows.

In Chapter 2, relevant background on 802.11 Wireless LAN networks is presented together with related work on radio resource management in cellular networks and WLAN networks. The proposed WLAN radio resource management framework from this research is briefly introduced at the end.
The detailed proposal for this research is introduced in Chapter 3. The intelligent agent based distributed control was selected as the first approach to optimise WLAN networks. The prototype sectored semi-smart antenna and control mechanism is briefly explained, followed by proposed solutions for locating WLAN mobile clients and monitoring the real time traffic load.

In Chapter 4, the intelligent agent based algorithm is discussed in detail. Some background information on intelligent and multiple agents (the software used in agent cooperation) is introduced. The agent based negotiation framework is discussed in detail and the novel 802.11 WLAN simulator is then explained and the validation of the simulator is discussed. At the end of the chapter, agent-based simulation results are analysed and a significant improvement is discovered when intelligent agents are used in a congested (hot spot) scenario.

As the most WLANs work in an indoor multi-path rich environment, some site-specific characteristics must be taken into consideration. In Chapter 5, the agent-based approach is further improved to work in constrained environment scenarios. The new method is used to simulate radio propagation in a real environment, like an airport lounge. The antenna pattern from the semi-smart antenna is also applied to the simulator instead of an ideal pattern. Simulations are performed against an imaginary large hall with walls and obstacles and results show that even with four access points, the network can still improve its capacity up to 30%.

Chapter 6 introduces a new centralised global optimisation algorithm. This algorithm use iterations to find the optimisation solutions when semi-smart antennas are used in WLANs. Simulation results show encouraging improvements compared with the conventional network and the agent-based approach. The algorithm works well because it can co-operatively shape the coverage of all access points towards the hot spot with built-in directional accuracy.

Finally the thesis is concluded in Chapter 7.
Chapter 2  Background

2.1  Introduction

Over the last few years, wireless data communications using the IEEE 802.11 standard has become widespread. Also known as Wireless Local Area Network (Wireless LAN or WLAN) [2], it is designed to replace Ethernet wired local area networks and offers the advantages of mobility and ease and speed of deployment. In particular, public wireless hotspots have become very popular at locations like airports, hotels and some coffee shops.

The initial 802.11 standard was finalised in 1997 but it was only recently that the 802.11 based wireless LAN products have achieved wide penetration. As IEEE 802.11 WLANs operate in the ISM license free radio band [25], it has the great advantage compared with terrestrial mobile networks which require networks operators to pay significant amounts of money to Government in order to use the radio spectrum. With 802.11 WLAN, anybody can setup their own WLAN networks without a licence, which is why so many WLAN hotspots have been set up all over the world. Now, users can easily use their WLAN enabled mobile phones, PDAs and laptops to access the Internet at many locations in most countries. For example, the organisation Boingo that allows access to WLAN hotspots across the world1 lists 10546 hotspots in the UK alone.

A competitor to the IEEE 802.11 standard is HIPERLAN (High Performance Radio LAN) [26], a European alternative (defined by European Telecommunication Standards Institute). The goal of the HIPERLAN was a data rate higher than that in 802.11. A later HIPERLAN standard (version 2 or HIPERLAN/2 [27]) was designed as a fast wireless connection for many kinds of networks using the 5 GHz band with a data rate of up to 54 Mbit/s. Although HIPERLAN claims to have many superior functions [28] the IEEE 802.11 standards have already occupied the market that HIPERLAN was designed for, so it is no longer regarded as a viable competitor.

Another important development in wireless broadband technology is the IEEE 802.16 WiMAX standard [29]. Work on the standard was started by the IEEE Standards Board

---

1 www.boingo.com
in 1999, with the aim of preparing formal specifications for the global deployment of broadband Wireless Metropolitan Area Networks (WirelessMAN). IEEE 802.16 was originally designed for high-speed wireless data connection for long distance fixed stations, e.g. providing a wireless alternative to cable and DSL for last mile broadband access; connecting Wi-Fi hotspots with each other and to the Internet. The latest developments of 802.16, however, have concentrated on mobility of subscribers, where it could be a competitive technology for Wi-Fi and 3G.

This research concentrates on the 802.11 family so alternatives will not be considered further, but it should be borne in mind that the principle behind the research described here could also be applied to other standards like 802.16, but the details would differ.

2.2 The IEEE 802.11 standard

The IEEE 802.11 standards cover the lowest 2 layers of the seven-layer OSI reference model [30]: Medium Access Control (MAC) layer and Physical (PHY) layer.

IEEE 802.11 networks consist of the following major physical components [31]:

- the Distributed System (DS), which is the backbone network for connected access points to forward frames to their destinations (Ethernet is most commonly used technology for the DS);
- Access Points (APs), the devices that perform the wireless-to-wired bridging function;
- Stations (STAs), which are typically battery-operated laptop or handheld devices.

The basic building block of an 802.11 network is the basic service set (BSS) [32], which is simply a group of stations that communicate with each other. 802.11 devices can operate in two different modes: ad hoc mode and infrastructure mode (as shown in Figure 2.1).

- In the ad hoc mode (formal name independent BSS, or IBSS), stations communicate directly with each other and thus must be within direct communication range. Typically, AD Hoc networks are composed of a small number of stations set up for a specific purpose and for a short period of time.
- The more common mode is called the infrastructure where all STAs associate
with an AP to obtain network services and then all communications from that STA are routed through that AP.

![Diagram of Ad Hoc and Infrastructure modes]

**Figure 2.1 Ad Hoc and Infrastructure modes**

Association is the process by which a mobile station joins an 802.11 network. In 802.11 standards, mobile stations always initiate the association process and the AP may grant or deny this association request. Also, a mobile station may only associate with one AP.

Mobility while communicating is one motivation for deploying an 802.11 network: stations can move and be used while connected to the network. Multiple APs can cover a large area in an overlapping manner in order to provide continuous coverage. This is somewhat similar to mobile cellular networks and mobile stations can move from one BSS to another without losing connection. This function is somewhat similar to handover in cellular phones handoff, with the difference that 802.11 networks are packet based. In the following example, at the outset, denoted by t=1, the laptop with an 801.11 network card is sitting with AP1’s basic service area and is associated with AP1. When the laptop moves out of AP1’s BSS and into AP2’s at t=2, a BSS transition occurs. The mobile station found itself out of AP1’s BSS and enters AP2’s. it use the reassociation service to associate with AP2, which then start sending frames to the mobile station. In this scenario, AP2 needs to inform AP1 that the mobile station is now associated with AP2.
The reassociation or handoff is different from that in cellular networks. In 802.11, the mobile station plays the main role in activating a handoff. When operating, the mobile station monitors all signal strengths from different APs. Within an Extended Service Set (ESS), the 802.11 standard offers MAC layer mobility, but between ESSs, mobility needs upper layer support, such as Mobile IP [33].

The detailed procedure of 802.11 handoff can be divided into three steps (Figure 2.3):

**Step 1: Discovery:** a STA scans nearby APs to be candidates for transition, with either passive scan or active scan being used. With passive scan, the station does not send probe request frames (saving power); the station listens to beacons sent from all APs in every available air channel and all information is temporarily saved in order to select a best candidate AP. In active scan mode, the mobile station actively send out probe requests on every channel and waits a certain time for probe responses; this is much faster than passive scan mode as it does not have to wait for beacon frames and it is also the default mode for most stations.

**Step 2: Authentication:** for security reasons, if necessary, a station is required to authenticate itself to the new AP in order to access the network.

**Step 3: Reassociation:** the last transition step a station continuously monitors the signal quality from all APs and selects the best AP to perform the reconnection.
This BSS transition (roaming) requires the cooperation of APs, as when the laptop reassociates with AP2, AP2 must communicate with the old access point (AP1) to determine that a previous association did exist and poll any buffered frames from AP1 to the laptop. Also, the AP1 must terminate its association with the laptop as mobile stations are allowed to associate with only one access point at any given time.

However the original 802.11 standards did not define the communications between the APs during BSS transitions, so that different vendors have developed their own standard to support mobility. For example, BreezeCom claims their 802.11 product line provides a roaming mechanism that allows stations to roam at speeds of 60 km/h without losing or duplicating packets [34].

In order to standardise inter-AP communications, the 802.11 working group drafted a standard called 802.11f [35] (or Inter-Access Point Protocol (IAPP)) as a recommendation for communications between multi-vendor systems. As the requirement for multimedia services like VoIP becomes more and more demanding for WLAN, with users expecting to be more and more while moving, a new standard called 802.11r [36] has emerged to specify fast BSS transitions. The new standard will permit connectivity from moving vehicles, with fast handoffs from one AP to another managed in a seamless manner. Compared with 802.11f, it will not only support data applications, but also applications like voice and video. The 802.11r standard is currently under development.
2.2.1 802.11 MAC

Fundamental to understanding the characteristics of 802.11 is the MAC and its multiple access mechanism. The 802.11 MAC controls the transmission of the user data into the air, providing core framing operations and interaction with a wired network backbone.

The 801.11 MAC adapts Ethernet-style networking into radio links. Like Ethernet, 802.11 uses a Carrier Sense Multiple Access (CSMA) scheme to control access to the transmission medium. In CSMA, a station desiring to transmit senses the medium; if the medium is busy (i.e., some other station is transmitting) then the station will defer its transmission to a later time, but if the medium is sensed free then the station is allowed to transmit. This protocol works well when the medium is not heavily loaded, but there are chances that stations transmit at the same time (collision) when two or more stations have sensed that the medium is free and have decided to transmit at the same time. These collisions waste valuable transmission capacity and must be identified. In Ethernet, collision detection (CSMA/CD) is deployed, so that if a station detects collision when it is sending a frame, it immediately stops sending and defers accessing the medium. But this detection mechanism cannot be used on a WLAN environment, because:

- Implementing a collision detection mechanism would require the implementation of a full duplex radio, capable of transmitting and receiving at once (an approach that would increase the price significantly).
- In a wireless environment all stations may not hear each other (which is the basic assumption of the collision detection scheme). Also, the fact that a station willing to transmit senses the medium free, does not necessarily mean that the medium is free around the receiver area. This situation is also called the “hidden node” problem (and will be explained later).

On a wired Ethernet, a station that transmit a frame, will assume that the destination receives it correctly. Radio links are different because noise, interference and multi-path fading may lead to situations in which frames cannot be transmitted to the destination. 802.11 uses a collision avoidance (CSMA/CA) mechanism together with a positive acknowledge scheme in which all transmitted frames must be acknowledged and without acknowledgement, the frame is considered lost and must be re-transmitted.
As mentioned before, in a wireless network, each node may not be able to communicate with every other node, as in Figure 2.4 (the "hidden node" situation).

![Figure 2.4: Nodes 1 and 3 are "hidden"]

In this figure, node 2 can communicate with nodes 1 and 3, but nodes 1 and 3 can not communicate directly. From the perspective of node 1, node 3 is a “hidden” node. Without hearing each other, nodes 1 and 3 could easily transmit simultaneously, thus rendering node 2 unable to make sense of anything. Further more, nodes 1 and 3 would not have any indication of the error because the collision was local to node 2. To prevent collisions, 802.11 standards allow stations to use Request to Send (RTS) and Clear to Send (CTS) signals to clear out an area. For example, in Figure 2.4, node 1 has a frame to send and it initiates the process by sending an RTS frame. The RTS frame can reserve the radio link for transmission as well as silence any stations that hear it. When node 2 receives an RTS, it responds with a CTS, which, like the RTS frame, silences stations in the immediate vicinity (i.e. node 3). When the RTS/CTS exchange is finished, node 1 can send its frame without worry of collisions from any hidden nodes. Thus hidden nodes (node 3) beyond the range of the sending station (node 1) are informed by the CTS from the receiver. As the RTS/CTS procedure consumes a fair amount of capacity, it is used in high-capacity environments with significant contention of transmission. For a low capacity environment, it is not necessary. The RTS/CTS is tuneable by adjusting the RTS threshold, where RTS/CTS exchange is only performed when the frame size is larger than the threshold.

Access to the wireless medium is controlled by coordination functions: CSMA/CA access is provided by the Distributed Coordination Function (DCF) and if contention-free access is required, the Point Coordination Function (PCF) can be used.
Contention-free services are only used in infrastructure mode.

A carrier-sensing function is used to determine whether the medium is available. Physical carrier-sensing and virtual carrier-sensing functions are used in 802.11:

- Physical carrier-sensing functions are provided by physical layer, but it is difficult to build physical carrier-sensing hardware as that would require the implementation of a full duplex radio, capable of transmitting and receiving at once (increasing the cost significantly) and, furthermore, hidden nodes make physical carrier sensing not able to provide all the necessary information.
- As a result, virtual carrier-sensing is used in addition to physical carrier-sensing. The Network Allocation Vector (NAV) is utilised for this function and the RTS/CTS example illustrated above is a good example of this function. In Figure 2.5, the sender station transmits RTS in order to reserve the medium; this frame contains source, destination and the duration of the following transaction (i.e. the time to transmit the data frame and respective ACK). The receiver station then responds with a CTS, which includes the same duration information. All other stations receiving either RTS or CTS will set the NAV indicator for the indicated duration and assume the medium is busy.

![Figure 2.5 Using NAV for virtual carrier-sensing](image)

Distributed Coordination Function (DCF) [37]

DCF is the basic standard CSMA/CA access function. Stations needing to transmit frames must sense the radio medium first and make sure it is clear before sending. To avoid collision, stations use a random back-off after each frame. RTS/CTS can be used
to further reduce the possibility of collisions.

In DCF, a station senses the channel to determine if another station is transmitting before initiating a transmission. The station proceeds with its transmission if the medium is determined to be idle for an interval that exceeds the Distributed Inter-Frame Space (DIFS). If the medium is busy, the transmission is deferred until the end of the ongoing transmission. A random interval, also referred to as the back-off interval, is then selected and is used to initialise the back off timer, or the so-called Contention Window (CW); the duration of this timer is determined as a multiple of a slot time, e.g. 9 µs in 802.11a. The backoff timer is decremented only when the medium is idle; it is frozen when the medium is busy. After a busy period the backoff timer resumes counting down after sensing the channel as being idle again. In this manner, stations that deferred from channel access because their random backoff time was larger than the backoff time of other stations, are given a higher priority when they resume their transmission attempt. After each successful transmission, a new random backoff is performed by the station that has successfully transmitted. If the sending station has not been acknowledged (i.e. does not receive an ACK), it increases its CW size. After any unsuccessful transmission attempt, another backoff is performed with a doubled size for the CW. This reduces the collision probability when there are multiple stations attempting to access the channel.

To limit the probability of long frames colliding and being transmitted more than once, data frames may also be fragmented. With fragmentation a large frame can be divided into smaller ones, which can then be transmitted sequentially as individually acknowledged data frames. The benefit of fragmentation is, if there are failed transmissions, the error is detected earlier and there is less data to re-transmit. The obvious drawback is the increased overhead. The basic DCF procedure is illustrated in Figure 2.6.
Interframe Spacing [37]

Interframe spacing plays an important role in coordinating access to the transmission medium. A varying interframe spacing creates different priority levels for different types of traffic. As shown in Figure 2.7, the SIFS is the shortest and used for the highest priority transmissions (such as RTS/CTS frames and ACK). High-priority transmissions can begin once the SIFS has elapsed, this will ensure conjunctive events (e.g. RTS-CTS-Frame-ACK) are not interrupted by other low priority frames. PIFS (PCF interframe space), is used for PCF during contention-free operation. Stations with data to transmit in the contention-free period can transmit after the PIFS has elapsed. It is shorter than DIFS because PCF traffic is higher priority than DCF traffic.

Point Coordination Function (PCF) [38]

PCF provides contention-free services. A special station (i.e. the AP) is used to schedule all stations to transmit frames in a coordinated manner. APs send “beacon” frames at regular intervals (usually every 0.1 second). Between these beacon frames, PCF defines
two periods: the Contention Free Period (CFP) and the Contention Period (CP). In the CP DCF is used, but in the CFP, the AP sends Contention Free-Poll (CF-Poll) frames to each station, one at a time (in a round-robin fashion for all stations), to give them the right to transmit a frame. This allows for a better management of the QoS, such as throughput and delay. This is beneficial for time-sensitive applications like VoIP. Unfortunately, PCF does not define classes of traffic and also, PCF is not widely commercially implemented so it will not be discussed in detail here.

IEEE 802.11e standard

With the growth in delay-sensitive services on the Internet (services such as VoIP, and video conferencing) QoS requirements have become more important than before. In order to offer Quality of Service (QoS) guarantees for 802.11 WLAN, a new standard (802.11e [4]) has recently emerged to enhance the MAC function by enhancing the DCF and the PCF with a new co-ordination function: the Hybrid Coordination Function (HCF). With the HCF, two channel access methods similar to legacy 802.11 MAC are defined: the HCF Controlled Channel Access (HCCA) and Enhanced DCF Channel Access (EDCA), both of which define Traffic Classes (TC). For example, VoIP would be assigned to a higher priority TC than email. In EDCA, a high priority TC has a higher probability of being sent than a low priority TC and so it waits less time.

As 802.11e was only standardised in late 2005, there are currently almost no commercial products available so it is not considered specifically in this research. However, any form of QoS mechanism must have sufficient resources otherwise low priority traffic will be starved of resources – and this research is about providing resource in the right place.

2.2.2 IEEE 802.11 Physical layer

The second major component (and most developed one) is the physical layer (PHY). The original 802.11 standards ratified in 1997 had three PHYs, two of these PHYs provide communications in the 2.4 GHz Industrial Scientific and Medical (ISM) bands using Direct Sequence (DS) and Frequency Hopped (FH) Spread Spectrum techniques; the third PHY is for infra-red (IR) links. The advantage of using the ISM band is that it is unlicensed spectrum so is free to use.
Data rates of up to 2 Mbit/s can be obtained with each of the PHYs and, because of these low data rates, very few products were produced using these initial 802.11 standards.

More PHY standards have subsequently been released: 802.11b, 802.11a and 802.11g. Both 802.11b and 802.11a were ratified in 1999, but the real widespread adoption of the 802.11 networks only occurred after products based on 802.11b appeared.

IEEE 802.11b [2]

The 802.11b variant operates in the 2.4 GHz ISM band and supports raw data rates up to 11 Mbit/s and offered a significant improvement over the original maximum speed of 2 Mbit/s. 802.11b uses an extension of the DS modulation technique defined in the original standard: instead of the original Differential Binary Phase Shift Keying (DBPSK) for 1 Mbit/s and Differential Quadrature Phase Shift Keying (DQPSK) for 2 Mbit/s, it uses Complementary Code Keying (CCK) as its modulation technique, which is a variation of CDMA. It added 5.5 and 11 Mbit/s to the 1 and 2 Mbit/s data rate and led to the rapid introduction of WLAN products, rapidly becoming the definitive wireless LAN technology.

The 2.4 GHz ISM is divided into 14 channels for 802.11 to operate (2.412-2.484 GHz) with 5MHz bandwidth in each channels. Channels 1 to 11 can be used in most parts of the world (13 channels for Europe and 14 for Japan). However, because of the nature of the 802.11 DS, most of the energy within one channel is spread across a 22 MHz band. To prevent interference from networks operating on adjacent channels, 802.11 DS equipment should be separated by a frequency band of at least 22 MHz between channel centre frequencies; this leads to channels 1, 6 and 11 being used as probably the most commonly used non-overlapping channels [39] (they are separated by 25 MHz). The channel allocations of 802.11b are shown in Figure 2.8.
The 802.11a standard was ratified at the same time as 802.11b in 1999. It operates at the much higher frequency of 5 GHz (UNII band in the U.S.), supporting raw data rates up to 54 Mbit/s as well as 48, 36, 24, 18, 12, 9 and 6 Mbit/s if required (depending on signal and noise power). The higher data rate was achieved by a different modulation technology, Orthogonal Frequency-Division Multiplexing (OFDM), which reduces multipath effects and increases spectral efficiency. Since the 2.4 GHz band is heavily used (by cordless phones, Bluetooth and even Microwave ovens), the 5 GHz band gives 802.11a the advantage of less interference. However, this high carrier frequency also brings disadvantages as it restricts the use of 802.11a to almost line of sight (higher frequencies are more likely to be obstructed) and short range, which requires the use of more APs.

Since the introduction of 802.11g (below) 802.11a is rarely seen so will not be considered further.

802.11g [40]

802.11g was the third extension to the modulation standard extension to be ratified (in June 2003), working at the same radio band and with the same channels as 802.11b, but supporting a raw data rate up to 54 Mbit/s. The 802.11g hardware is also backward compatible with 802.11b hardware.

802.11g uses OFDM for data rates of 6, 9, 12, 18, 24, 36, 48 and 54 Mbit/s and reverts to CCK for 5.5 and 11 Mbit/s (like 802.11b) and DBPSK/DQPSK for 1 and 2
At the time of writing this thesis, 802.11g systems have taken the dominant market position with dual band (802.11 b/g) being the normal offering and tri-band (802.11 a/b/g) products also being quite common.

**802.11n [41]**

IEEE has announced the development of a new amendment to the 802.11 (802.11n) standard that can reach a theoretical data rate of 540 Mbit/s. The new standard builds upon previous 802.11 standards by adding MIMO (Multiple-Input Multiple-Output), which uses multiple transmitter and receiver antennas to give increased data throughput through spatial multiplexing and increased range by exploiting the spatial diversity. The new standard is currently under development and according to the IEEE 802.11 Working Group project timelines, the 802.11n standard is not due for final approval until April 2008.

To support very noisy environments as well as extended range, 802.11 WLANs use *dynamic rate shifting*, allowing data rates to be automatically adjusted to compensate for the changing nature of the radio channel. Ideally, users connect at the full data rate, but when the user moves beyond the optimal range for full speed operation, or if substantial interference is present, the STA falls back to a lower speed; if the STA moves back into a better radio environment, the connection will automatically speed up again. Rate shifting is a physical-layer mechanism transparent to the user and the upper layers of the protocol stack.

The 802.11 standard itself does not define the rate switching algorithm, but only defines the rules. These rules, include how management frames, multicast and unicast frames should be transmitted and to how to make sure both the transmitter and receiver stations support the data rate. The switching algorithm is achieved differently by different vendors based on the bit error rate or signal to noise ratio.

Fall-back rates for 802.11b are 5.5, 2, and 1 Mbit/s and for 802.11a (nominally 54 Mbit/s) 48, 36 and intermediate rates down to 6 Mbit/s.

The 802.11g standard has the same propagation characteristic as 802.11b, because it transmits in the identical 2.4 GHz band, but the 5 GHz radio signals of 802.11a do not propagate as well so the range of 802.11a devices is less than the 802.11b or 802.11g product range.
Figure 2.9: Expected 802.11a, 802.11b and 802.11g data rates at varying distance from access point [42]

Throughput is not the same as data rate for networking systems because of factors such as overhead: the maximum throughput of 802.11b is only 47% of 11 Mbits/s and for 802.11a and 802.11b a net throughput of 55% of 54 Mbits/s has been observed [42].

2.3 RRM for Wireless Networks.

Unlike 802.11 wireless networks, mobile network operators may have to pay for spectrum and also base stations are considerably more expensive than WLAN APs, so that research on getting the most from the radio resource available has been much more widespread for cellular networks than that for WLANs. In that research the distribution of mobile users and the topology has been an important factor [43, 44].

There are several ways to improve the capacity of a cellular system without adversely affecting the quality of service:

(a) decreasing the frequency reuse factor;
(b) reducing interference; and
(c) balancing the load among base stations.

In a TDM-based network, e.g. GSM, different channel allocation schemes maybe used
to improve system capacity (approaches (a) and (b)). The base station can adaptively select frequencies based on interference for each user in the cell [13], [45]. For CDMA-based networks, as all cells use the same frequency band, channel allocation is not applicable and other approaches have to be used.

To apply approach (c) in cellular networks, techniques such as cell splitting, cell sectorisation and cell breathing are used. Cell splitting creates more coverage and capacity in a wireless system by having more than one cell site cover a particular geographical area, each cell site covering a smaller area in the high density area [46]. Adaptive cell sectorisation is another technique that partitions a single cell into multiple adaptive sectors according to traffic distributions [47-49]. Cell breathing is another well-known method that adaptively adjusts a base station’s pilot power according to local traffic load [50, 51].

Another approach to the overall problem is using smart antenna technology and research in [15] has shown that using smart antennas to control radio resources can offer significant improvements in radio resource management. In this research, the aim is to investigate whether that line of improvement can be applied to WLANs.

2.3.1 Smart antenna research

A smart antenna system combines multiple antenna elements with a signal processing capability to optimise its radiation and/or reception pattern automatically in response to the signal environment [52]. Usually, it consists of a number of radiating elements, a combining or dividing network and a control unit. The control unit is where the intelligence lies. A considerable amount of work has been published in the literature on smart antenna systems, for example [53-57]. The radiation pattern can be regarded in a broader sense as cell size and shape, and the work in, for example, [17] has already shown that changing both cell size and shape is feasible and can increase system capacity.

Smart antenna systems are usually categorised as either switched-beam or adaptive-array systems. Both systems attempt to increase gain in the direction of users but only the adaptive-array system offers optimal gain [58].
Figure 2.10: Illustration of smart antenna types

- **Switched beam**: in this approach, the smart antenna has a finite number of fixed, predefined patterns and the control system switches from one beam to another as the mobile moves throughout the sector.

- **Adaptive array antennas**: here the number of patterns is not chosen in advance, but are adjusted in real time. The adaptive system can locate and track signals to dynamically minimise interference and it is possible to change the size and shape dynamically.

The work in the EU project SHUFFLE [59] tested the use of an adaptive antenna array for a 3G base station, effectively adding an extra control parameter (radiation shape) that can be used to optimise the performance of the network. A broad change in cell coverage was realised by using adaptive arrays with a fairly small number of elements. This type of antenna was named a *semi-smart antenna* and the concept of using such antennas for 3G network was implemented in [18, 19, 59]. Figure 2.11 shows the four-element array used for testing in that project.

---

2 Source [52] ArrayComm, "Smart Antenna System," http://www.iec.org/online/tutorials/smart_ant/, 2000. Figure 6 & 7
Figure 2.11: Prototype semi-smart antenna\(^3\) and its beamforming.

With only 4 elements, it is still possible to change the shape and size of the radiation pattern sufficiently: the principle is illustrated in Figure 2.11. It shows a three-sector antenna and the radiation of one sector (4 elements). In the left pattern, equal excitation is used on each antenna element and if all three sectors use this equal excitation, a nominally circular radiation pattern is produced for the cell. The right pattern shows varied excitation for each element of the sector, and those shapes may vary sector by sector so that the base station can produce a flexible coverage shape.

The more elements and sectors that are used, the more degrees of freedom there will be and hence a greater variety of shapes; however, in a real system more elements would mean a higher cost.

Although the benefits of using “fully” smart antenna systems are many, there are drawbacks and cost factors. Antenna beam forming is a computationally intensive process, especially if a fully adaptive-array is used. The complexity and cost of the

\(^3\) Constructed from four conventional base station antennas placed side by side (with built-in remote electrical tilting and phase shifting – a proprietary system from Alan Dick and Company)
adaptive beam-former is still seen as a major disadvantage for the fully “adaptive smart antenna” system.

In this research, the semi-smart antenna concept is used, however, as the cost of a WLAN system is very much lower than a cellular system, a simplified version (with low cost) of the semi-smart system is used. The concept and design of such a low-cost semi-smart system is discussed in Chapter 3 section 3.1 and 3.2.

2.4 Extending Radio Resource Management (RRM) functions to 802.11 networks.

2.4.1 Introduction

The 802.11 WLAN was initially designed to be used with a single AP that would cover all the required geographical area. Later, installations were set up with APs placed close enough to each other to offer a continuous coverage over a wider area. However, an improperly designed WLAN can easily leave coverage holes and significantly lower data rates for users, especially when many users are using a single AP heavy. These users will experience lower data throughput and longer delay due to congestion occurring in the air channel. In order to tackle the problem, a common WLAN deployment practice [60] has been designed and used.

2.4.2 The common WLAN deployment practice

WLAN deployment is inherently complicated, especially when used indoors, because of the different RF characteristics of different buildings, and hence the different radio propagation and interference issues that need to be overcome. The dynamic nature of the wireless channel makes it even more difficult to provide uninterrupted connectivity.

In terms of radio resource management, the issues that need to be addressed are:

(a) coverage – to provide network connectivity to all desired locations; and
(b) capacity - provide sufficient bandwidth to satisfy users’ needs.

The common industrial practice to manage the coverage and capacity of WLANs starts by designating the coverage areas. Next, through capacity and coverage calculations,
the required number of access points per coverage area is determined and their power-levels calculated to meet the required cell dimensions. Access points are then assigned frequency channels keeping in mind the capacity and co-channel interference constraints. Finally, after installation, a physical site survey is carried out to ensure the desired coverage and capacity are obtained, the deployment plan being fine-tuned to meet the network requirements. This process is illustrated in Figure 2.12.

![Figure 2.12 Manual WLAN deployment process](image)

The steps in this flow chart are as follows:

**Designating Coverage Area:** deciding the physical areas that needs coverage.

**Capacity Planning:** determining the bandwidth requirements of each area. The bandwidth requirements can be derived from estimating user density and their average bandwidth requirements. This step only approximates the long-term bandwidth available to users.

---

4 The diagram is derived from Fig. 2. in [60] A. Raniwala and T.-c. Chiueh, "Coverage and Capacity Issues in Enterprise. Wireless LAN Deployment," Stony Brook University 2004.
Coverage Planning: calculating the number of APs required to ensure overall coverage. In low capacity regions, the number of APs is lower than high capacity regions.

Preliminary AP Positioning and Power Selection: in this step, APs are physically placed across the facility. Without the use of a sophisticated software tool, this step is generally based on practical deployment experience. The AP power-level is initially set to the maximum possible and more APs are added if coverage holes are discovered. Subsequently, the power level of the APs is reduced if there is too much overlap between adjacent cells, but some amount of overlap has to be maintained to ensure continuous coverage to mobile clients.

Channel Allocation then has to take place to reduce interference among APs that are within the interference range of each other. A typical strategy is to use a direct-sequence channel layout as shown in Figure 2.13. In this layout a pattern is assigned to a cluster of cells, and is repeated across clusters, this example allocation is based on 3 non-overlapping channels of 802.11b/g.

![Figure 2.13: Frequency reuse pattern for 801.11b/g](image)

Physical RF Site Survey: once the provisional APs have been deployed, the most painstaking part of the deployment beings – the RF site survey. The purpose of the step is to ensure that the predicted coverage and capacity is being obtained. Some software tools are available for assisting this step [61] by using a laptop or a PDA. The network designer visits several pre-defined locations and measure various configuration/performance parameters.


### 2.4.3 Limitations of the common practice and various RRM schemes

In some scenarios, the distribution of users changes dynamically: for example, in an airport, users tend to gather at different departure gates at different times. To cope with these changing demand patterns, more APs need to be purchased to be placed in areas where there might be high demand at some time. Moreover, as well as increasing the costs, such an approach also leads to interference with the limited number of channels available.

Even properly deployed WLANs can not be maximally optimised in these scenarios. Recent studies [62, 63] on operational 802.11 WLANs have shown that the traffic load is often un-evenly distributed among the APs. In WLANs, by default, a user terminal scans all available channels to detect its nearby APs and associates itself with an AP that has the strongest received signal power, irrespective of the load on the APs. As users are generally not evenly distributed, some APs tend to suffer from heavy load while their adjacent APs may carry only light load. Such load imbalance among APs is undesirable as it prevents the network from fully utilising the network capacity.

Currently the IEEE 802.11 standard does not provide any standard method to resolve the load imbalance. To overcome this deficiency, several researches have been carried out (before or during the same time of this research) with the aim of solving the un-balanced traffic problem. That research can be summarised into two categories.

(a) Association-Time Load Balancing

Most of these methods take the approach of directly controlling the user-AP association by deploying proprietary client software or specially-designed WLAN cards in the user computers. For example, some vendors [64] claimed they have incorporated certain load-balancing features in their device drivers, AP firmware, and WLAN cards. In these proprietary solutions, the APs broadcast their load levels to users via modified beacon messages, and each user chooses the least-loaded AP.

Several studies [7-11] have proposed a variety of association metrics instead of using the received signal power (RSSI: Received Signal Strength Indication, that is commonly used in 802.11) of the AP as the sole association criterion. These metrics typically take into account such factors for an AP as the number of users currently associated, the
mean RSSI of users currently associated, and the bandwidth that a new user can get if it does associates (e.g. [7, 8]). Balachandran et al. [9] proposed to associate a user with an AP that can provide a minimal required bandwidth by the user and if there is more than one suitable AP, the one with the strongest RSSI is selected. Velayos et al. [10] suggested a distributed load balancing scheme where the load of an AP is defined as the aggregated downlink and uplink traffic through the AP. In [11], Kumar et al. discussed an association scheme based on the concept of proportional fairness to balance between throughput and fairness (here fairness refer to the STAs associating to APs with certain bandwidth fairly).

(b) Dynamic Load Balancing

The association-time load balancing technique above becomes ineffective once the traffic pattern of associated clients changes. A more dynamic scheme observes load imbalance and re-associates clients to less loaded APs. The key mechanism required for such post-association load balancing is the ability to forcibly switch a client from one AP to another based on traffic load information. This mechanism has to be triggered by an explicit notification from the AP rather than from a client-side signal strength measurement. In [6], Tsai and Lien proposed to re-associate users when the total load exceeds a certain threshold or the bandwidth allocated to users drops below a certain threshold. Bejerano et al. [12] used an on-line scheme that periodically optimises the user-AP association.

2.5 Summary

In this Chapter the basic principles of 802.11 WLANs have been introduced and possible load-balancing schemes summarised. However, these association-time load balancing and dynamic load balancing schemes described above are all based on modifying the current association mechanism of 802.11 and suffer the disadvantage that they do not readily cope with changing geographical distributions of users,

In the next Chapter, a new dynamic WLAN load balancing scheme is proposed that does not change the current association approach of the 802.11 standard. The aim is to re-balance the network’s traffic load by changing the coverage patterns of the APs without modifying the current 802.11 standard. In addition, this solution is fully automatic, being accomplished using intelligent algorithms in a distributed or
centralised control. Traffic hotspots can be automatically balanced and users in those areas will get a much higher data rate compared with the traditional network setup.
Chapter 3  New RRM framework for 802.11 WLANs

This Chapter describes the novel approach to radio resource management for WLANs that forms the research presented in this thesis. The aim is to maximally utilise radio resources of a WLAN network in congested situations, offering mobile users improved data rates without modifying current 802.11 standards.

Although the concept is based on experience [15-17, 19, 20] of load balancing for wideband CDMA systems, this research is the first to apply that general approach to WLANs, and for indoor environments. This is a very different problem from that in [19].

In this research a distributed and later a centralised architecture are used to manage the radio resources. The distributed architecture is proposed first as it is more suitable for WLANs (since there is no central control structure); it also avoids the drawbacks in scalability, reliability, efficiency, and flexibility inherent in centralised architectures.

The distributed approach taken here is based on software agents, autonomous software entities that have their own state, behaviour, thread of control and the ability to interact and communicate with other entities [65]. The term agent has also been used for many years [66-69] in the field of distributed computing where it refers to specific (client or server) entities used in solving specific tasks in a distributed computing system.

3.1  A semi-smart sectored antenna for 802.11 AP.

In a cellular system, radio resources are reused after a certain distance. The whole area is divided up into a number of small cells, with one base station giving radio coverage for each cell. In a conventional network, the power control is fixed so that each base station’s radio coverage is static (neglecting the “cell breathing” of CDMA). In previous research [15-17, 19, 20], Du has developed a dynamic distributed agent-controlled system to change the coverage patterns in CDMA networks according to the
geographical traffic load so that the radiation pattern of cells is based upon the offered traffic load. Capacity in a heavily loaded cell can be increased by contracting the antenna pattern around the source of peak traffic and expanding adjacent antenna pattern to fill in the coverage loss. The idea of agent based distributed control is extended to WLAN in this research. Du later proposed a centralised control algorithm called the “Bubble Oscillation Algorithm” [18] again for WCDMA traffic balancing. The algorithm mimics a two dimensional coverage of air bubbles that will oscillate and tend to cover any gaps between the bubbles. For WCDMA networks, the local coverage scheme is treated as an air bubble, the local traffic load is treated as the air within the bubble, and un-served traffic is treated as a vacuum between adjacent bubbles. The process of geographic load balancing is performed by emulating the process of bubble oscillation, and the re-allocation of un-served traffic is fulfilled by the oscillations caused by the pressure difference between adjacent cells and attraction forces from temporary vacuums. In practice the algorithm could run in the Regional Network Controller in WCDMA networks.

In the latest research [70], Yao used learning algorithm to predict moving congestion patterns (moving hot spots) and then change antenna patterns in advance. So far, the studies only cover CDMA networks, and with out-door usage (ideal free space radio propagation model).

For WLAN systems, the “cell” concept can still be used to represent the AP layout and multiple APs can cover a wider area to allow mobility between APs. The most common antennas used in WLAN APs are omni-directional diverse antennas, which give a circular shape coverage area, although other patterns are available for specialised employment (such as along corridors). In order to achieve more controllable radio coverage patterns, this research assumes that every AP has a semi-smart antenna structured into four sectors of 90°. Each sector has multiple antenna elements and individual power control in order to achieve directional and controllable pattern, thus the combination of four sector antennas will give an AP the ability to produce flexible radio coverage patterns (Figure 3.1). The four-sector antenna patterns add to give the overall cell coverage pattern (Figure 3.2). In Figure 3.2 a), one face of the proposed

---

5 This is ignoring multi-sectored cells.
antenna radiates at different power levels (0 dBm and -10 dBm) (the radiation patterns are simulated using software); in b), the conceptual four faces’s radiation pattern is illustrated with one face changing its radiation power.

The choice of a four-sector antenna array is a compromise between performance and cost. More sectors will increase the flexibility of the coverage shaping but has the disadvantages that (i) more RF power controllers would be needed and (ii) each sector would need a narrower antenna pattern, so requiring physically bigger antennas. The cost of the proposed 4-sector antenna system is designed to be comparable with the cost of an AP.

### 3.2 New element array antenna for 802.11 WLAN

The detailed design of this antenna is outside the scope of this research; it was proposed by Parini in [71] but this work is the first to consider the control of such antennas in realistic WLAN deployment scenarios.

![Figure 3.1: The appearance and working diagram of the four-sector ceiling mounted antenna array.](image)
Figure 3.2: Coverage patterns of the antenna sectors with different power control

The four sector antenna pattern provides an overall coverage for each cell and by controlling the power to individual sector the overall shape can be changed. In the simple example shown in Figure 3.3, seven APs are laid out in a hexagonal shape to cover a large area as in cellular networks. If a the central AP becomes heavily loaded because of the number of users simultaneously using that cell, then the central AP can drop the power in its sectors and reduce the numbers of users associated to it. By this method, the cell traffic can be reduced, the surrounding APs then can deliberately increase the sectors towards the traffic hot spot and “pick up” those users previously associated to the central AP. From the system point of view, the users in the central hotspot are effectively distributed to several APs instead of crowded in a single AP.
Figure 3.3: Cooperative changing AP coverage patterns to reduce traffic hot spot.

The semi-smart antenna built is a microstrip patch antenna; this is the simplest configuration for a patch antenna, having the features of low cost, low profile, ease of configuration (e.g. specialised geometries), dual frequency capability, dual polarization, and being lightweight.

The final product would include the antenna being placed on the ceiling of a room (e.g. in an airport lounge). The antenna would be a pyramid shaped antenna as shown in Figure 3.4. The angle of the slanting edge of the antenna with the horizontal is approximately $45^\circ$ and the angle of the radiation towards the likely region of users is approximately $60^\circ$; this corresponds to a requirement of a 3dB beamwidth of $60^\circ$.

Figure 3.4: illustrates the angles required for coverage of users

---

The final patch antenna and the pyramid built looked like Figure 3.5 a) and b). A comparison between the theoretical (simulation) and measured radiations pattern are given in Figure 3.5 c).

Figure 3.5: a) the fabricated antenna; b) mounted on the four faces of the pyramid; and c) the theoretical (simulation) and measured radiations pattern.

The phase and amplitude of the excitation to each face may be changed by using a four-in-one power splitter, controlled by a microprocessor based control unit. Location of users

One of the main elements that affect the accuracy and efficiency of the intelligent RRM system is its ability to locate mobile users. With the knowledge of a user’s location, the Access Point Agent can effectively adjust the radiation pattern to cover those users in the most appropriate way. There are three main approaches for finding the location of a mobile device: angle of arrival, time difference of arrival and multiple signal strength computation. [73]

The aim of a positioning method is to provide the mobile device’s position in (x, y)
coordinates of a 2D or 3D map. To do this, there must be some reference points. The most popular method is the triangulation function (Figure 3.6) which knows the position of three base station and the corresponding distances to the mobile station. It is then very simple to calculate the mobile position relative to the base stations.

![Figure 3.6: triangulation to locate a mobile station](image)

In an “angle of arrival system” [74], the position is calculated via geometry. With the use of directional antennas or antenna arrays, the angle or bearing relative to points located at known positions is measured. The intersection of several measured direction pointers then yields the position value. The accuracy of this approach is limited by the possible directivity of the measuring aperture, by shadowing, and/or by multi-path reflections arriving from misleading directions.

The relative distances ($L_1$, $L_2$ and $L_3$) can be calculated by using one of the following methods:

**By time difference of arrival:**

This method calculates the distance $L$ by measuring the time difference of a signal pulse going from the base station to the mobile terminal (Figure 3.7). To do this accurately the signal must be exactly synchronised between the base station and mobile terminal and the easiest way to do this is to use the GPS system [75]. GPS receivers performed very well in open spaces, but due to severe multi-path...
propagation in cities and, additionally, the GPS signal is too weak to for indoor applications, they are not really suitable for WLAN locations since users are likely to be indoors. Multi-path is a real problem as receivers must be able to distinguish the time the direct signal is received from reflected signals. This approach has been used for locating WLAN terminals [76], but there are difficulties:

- To get location accuracy of the order of a few metres requires a high level of synchronisation accuracy.
- Most importantly, the location information produced is not accurate enough because of the variations in the received signal strength due to multi-path propagation.

*Figure 3.7: Time difference of signal pulses from AP to STA (Fig on page 6 of [76]).*

By calculating the signal strengths received from multiple APs:

The alternative way of calculating the distance from the base station to the terminal is to know the signal strength received.

The problem is that a propagation model does not work well indoors where signals will be reflected from walls, ceilings and objects in the rooms and will be absorbed by opaque objects like walls and human bodies. These effects will cause the received signal strength to be unpredictable and indeed highly variable.

To overcome this problem, advanced propagation models are required, or the actual field distribution in the area of interest has to be learned from measurements. The major advantage of a signal-strength-based approach is the fact that most modern radio modules already provide a received signal strength indicator (RSSI). Therefore, implementing a local-positioning system is more or less a software topic, and proprietary hardware is not required.
3.2.1 Ekahau position engine for locating WLAN users

The Ekahau Positioning Engine (EPE) [77] is a new, commercial, approach for locating WLAN users with a software solution. Rather than trying to model a complex indoor radio propagation model, the EPE uses a machine learning approach (the core estimation algorithms are Bayesian Networks [78]) to infer a model from a set of training data in order to obtain predictions. The characteristic of the approach is that it does not need any radio propagation model at all: site calibration involves taking signal samples at different known locations to construct a measured signal strength model based on the signal strengths from the different radio channels in use. The model is stored in the positioning engine. Ekahau clients will retrieve received signal strength values and return the values to the engine for location calculations.

Based on the sample points that created the measured signal strength model, the Ekahau positioning engine can estimate a client station’s approximate location giving an arbitrary combination of signal strengths. For a reasonably accurate prediction result (within a few metres) the client WLAN device needs to “hear” at least three access points, which should be operating in different non-overlapping channels. For example, Channel 1, 6 and 11 are the three recommended non-overlapping channels.

An evaluation of the Ekahau engine was carried out at QM by the author. The Ekahau Engine and Ekahau manager were installed on a laptop for testing. One screenshot of the tracking results is shown in Figure 3.8 (the current WLAN client: “Yapeng’s position” on the diagram). There are two access points, plus two access points on the floor below (they are not shown on the map as the locations of APs are not required for Ekahau); the whole area is an ideal test bed since it is an overlapping area with more than 3 access points being “seen” by a WLAN client almost anywhere. 90% of the location estimations are less than 4 metres away from the actual location. If switched to “accurate mode”, more accurate results (up to 2 metres) can be achieved after a slightly longer delay (about 2 seconds processing delay).
For locating wireless users in this research, the Ekahau approach is used, because this is a pure software approach and does not need proprietary hardware or modifications to the 802.11 standard. Also from experiments, the achieved accuracy (2 metres) is good enough for radio resource management purpose.

### 3.3 Monitoring traffic loads of WLAN Access Points

#### 3.3.1 Introduction

Monitoring real time traffic could be done most easily using special software using a proprietary protocol provided by the manufacturer to monitor the traffic load. However, access points do not generally have a public interface for third-party software and also different vendors supply their own software, which is generally not compatible with that from another vendor.

Davis [79] proposes a wireless traffic probe method for IEEE 802.11 WLANs capable of obtaining the traffic load of the network. An extra probe machine can passively sniff the wireless frames and analysis the contentions of DCF and derive the current load conditions of the network. However, this solution needs an extra machine to monitor each AP; in a large network composed of many APs this solution becomes
impracticable.

The method proposed here neither needs special proprietary software nor additional probes. Instead, monitoring some relative parameters of AP by SNMP is used.

SNMP (Simple Network Management Protocol) is an application layer protocol, part of the TCP/IP protocol suite, that facilitates the exchange of management information between network devices, enabling network administrators to manage network performance, find and solve network problems, and plan for network growth. Compared to proprietary protocols, SNMP is public and has freely available APIs.

A network management system contains two primary elements: a manager and agents. The manager is the console through which the network administrator performs network management functions. A typical manager is usually installed on a Network Management Station (NMS) and has the ability to query agents and get responses from the agents. Agents are the entities that interface to the actual device being managed. Switches and routers or network servers are examples of managed devices that contain managed objects. These managed objects might be hardware, configuration parameters or performance statistics that directly relate to the current operation of the device and are arranged in what is known as a virtual information database, called a Management Information Base (MIB).

A typical agent usually implements the full SNMP protocol, storing and retrieving management data as defined by the MIB. SNMP allows managers and agents to communicate for the purpose of accessing these objects [80]. There are four basic SNMP commands:

- **Read** command: used to monitor the network device.
- **Write** command: used to control the network device.
- **Trap** command: (operated by the network device) used to send reports to the NMS under some conditions.
- **Traversal** operation: used by the NMS to update route information in the route tables to be in line with the network device.

The MIB is a collection of definitions that define the properties of the managed object and every managed device keeps a database of values for each of the definitions
written in the MIB.

3.3.2 Investigating MIBs for 802.11 Access Point and Client.

Experiments have been carried out on an AP from Orinoco (Orinoco AP-200, the purchased AP for this research), which is for home and small business uses (at the bottom line of Orinoco Access Point family) and claims to have remote management ability through TFTP or SNMP. However, the MIBs that the software can support are MIB-II (RFC1213), Ethernet-like (RFC1398) and Bridge MIB (RFC1493). The 802.11 MIB is missing. More advanced AP models allow access to 802.11 MAC and PHY MIBs. This means that a safe line for this research is to assume that in most APs the set of MIBs found in the AP-200 would be available.

On the client side, because most of these devices are installed on personal computing devices, a local driver provides the normal way to manage the device’s behaviour. In order that the user resource agent can effectively perceive and control the 802.11 MAC and PHY layers, the resource agent program must have the ability to communicate with driver program.

The basic structure for radio resource management of WLAN is hence as shown in Figure 3.9.

![Figure 3.9: Structure for managing WLAN network](image)

This figure shows the architecture for managing the WLAN client and the AP: through the driver program of the client side and through SNMP on the AP side. One of the advantages of the above structure is ease of use: the AP agent would not necessarily...
reside in the AP hardware but on the server side by running an independent agent. Although SNMP heavily depends on the vendor’s MIB specification, and so would require some sort of middleware to interface with different types of AP, this approach does allow efficient monitoring and management of the underlying radio resource.

MIB-II is a standard MIB that defines the variables to manage the TCP/IP protocol stack. It has over 170 variables to support network layer management. Although the original purpose of design MIB was for fault management, it can also act as a traffic monitor through using counters.

The monitoring done this way is “passive”: it uses devices to watch the traffic and these are polled periodically and information is collected (in the case of SNMP devices the data is extracted from MIB) to access network performance and status. Active monitoring relies on the capability to inject test packets into the network, following them and measuring the service obtained from the network.

The objects in a MIB are distributed as a tree structure with digits being used to describe different branches of the tree. The leaves of the tree are objects defined by SNMP. The first level of the tree consists of 3 objects from the ISO and ITU-T [81]. There are 4 nodes below the ISO. The node representing the Internet object (1.3.6.1) is what is important here and most of the organisations and companies producing Internet products use this branch to identify the management information of the product. The second node below the Internet node is “mgmt” (management), marked by “1.3.6.1.2” and below that is the management information base (MIB-II), marked by “1.3.6.1.2.1”. Figure 3.10 shows this structure and the objects in the MIB.
If real-time traffic information was provided in the MIB it would be easy to monitor traffic load. However, there is no such information in the MIB-II of the managed AP-200 so a more generic approach has been implemented by periodically pulling the current number of received data with an OID object id “1.3.6.1.2.1.2.2.1.10” (named ifInOctets, indicate the total number of octect/Byte received/sent on the interface, which has been highlighted in green on the MIB-II tree). Hence, the current traffic throughput can be obtained by:

\[
\text{Speed}_{T_1-T_2}(\text{Mbit/s}) = \frac{\text{Counter}_{T_2} - \text{Counter}_{T_1}}{T_2 - T_1}
\]

(3.1)

Where \( \text{Counter}_{T_1} \) , \( \text{Counter}_{T_2} \) are the values of the retrieved current number of received data (in Mbits) at time periods \( T_1, T_2 \) respectively.

Furthermore, by accessing the 802.11 MIB, an AP agent can have in-depth knowledge of traffic conditions at its AP. This includes TransmittedFragmentCount and ReceivedFragmentCount (through which the MAC layer traffic condition can be estimated), FailedCount, which increments when the number of transmission attempts
has been exceeded (it is normal for this counter to rise with increasing load on a particular BSS so that it could also be used as an indication of congestion). Other relevant parameters are RetryCount, FCSErrorCount and FrameDuplicateCount.

### 3.4 Summary

In this chapter, the proposed mechanism to manage radio resource and improve system capacity for 802.11 WLAN is introduced. The key idea is to utilise low-cost semi-smart antennas that can change the radio propagation and balance congested APs when traffic hot spots are exist. The structure and control mechanism of the semi-smart antenna is discussed. Secondly, various WLAN positioning techniques are discussed and a pure software based solution is proposed to locate wireless clients in this research. At the end of the chapter, the mechanism of monitoring AP and STA’s traffic load is analysed and a non-proprietary software-based approach (over SNMP) is proposed and initial tests show the feasibility.
Chapter 4   Intelligent Agent Approach to RRM

The distributed intelligent agent control method for 802.11 RRM is introduced in this chapter. In this approach, all access points and the antennas are represented by software agents which can communicate and co-operate with each other using a standard agent communication language. The co-operation results are mutual changes of the radio coverage from access points in congested areas. Those heavy loaded APs are then relieved by handing over some of the mobile stations to adjacent APs. In this chapter, the intelligent agent approach and the associated software are introduced first and the agent-based negotiation framework is analysed in detail later. The system simulator and simulation results are discussed at the end.

4.1 Intelligent Agents

Intelligent agent based systems technology has generated a lot of interest in recent years because of its attractiveness as a new way of generating software for distributed systems [83]. In this context, an agent is a piece of software that can detect what is happening around it and make decisions to allow it to take appropriate action. A software agent should have some basic characteristics, i.e. autonomy, social ability, reactivity and pro-activeness [65].

“Autonomy: agents are able to operate without the direct intervention of humans or other agents, and have some kind of control over their actions and internal states.

Social ability: agents are able to interact with human or other agents, via some kind of communication language, to achieve a common goal or solve a problem together.

 Reactivity: agents perceive changes of their environment and respond with some timely proper actions.

Pro-activeness: agents have the ability to plan ahead and take the initiative to perform actions that will contribute to the goal achievement without waiting for
external instructions or only responding to events in the environment.”

Agents may also have some other properties, known as the stronger definition of agents, such as mobility, veracity, benevolence, rationality or learning/adaptation [84]. However, this work makes use of agents, it is not about of agent research per se so this section will not go into any details on these other properties.

There are two types of agents that have been used in telecommunications systems: intelligent agents and mobile agents. *Intelligent agents* are pieces of code that reside in a system and communicate with other agents using some sort of *agent communication language* (ACL); *mobile agents* are pieces of code that move around a system taking action at each place they reside. In general, intelligent agents have been used more in the literature [67, 85-87] because mobile agents could potentially be more of a threat: a rogue mobile agent moving around would be somewhat similar to a virus.

Intelligent agents need a structure to implement their behaviour in response to the conditions; several structures have been proposed to achieve this [88, 89].

Figure 4.1 (adapted from Fig. 1 of [90]) gives a simple example of how a utility function could be used to implement this decision making. The utility is a function that maps a state (or sequence of states) to a number that represents the agent’s “happiness”. The agent perceives the environment (i.e. it reads the values from the sensors/inputs attached to it) and checks its states and knowledge base to determine what it thinks the world is like now. Then it does some prediction and planning according to the environment and its own states, evaluates plans of action, and finally decides which to use and executes this plan. In this way, the agent can adapt and change continuously in response to changing conditions around it.

---

4.2 Multi-agent systems

Multi-agent systems (MAS) comprise a collection of agents (often a dynamic set). The agents reside somewhere in the network and are able to exchange messages with other agents located elsewhere in order to achieve their goals. Multi-agent systems can be used to implement complex problem-solving systems in the absence of centralized planning and such systems usually use the interaction of a large number of simple agents that are able to sense and change their environment locally.

Societies of simple agents are capable of complex problem solving even though they possess limited individual abilities.

The term multi-agent system is used for all types of systems consisting of multiple autonomous components showing the following characteristics:

- Each agent does not have sufficient capabilities itself to solve the problem.
- There is no global system control.
- Data is decentralized.
- Computation is asynchronous.
At a high level the multi-agent systems approach is intuitively simple and the co-operation between others is often analogous to co-operative problem solving between people - terms such as “request” and “confirm” illustrate that: software developers can draw on their own experience in solving problems co-operatively with others.

Combining the characteristics of intelligent agents and multi-agent systems gives greater intelligence and flexibility for controlling a large complex system. Resource management is a crucial aspect of any telecommunication system as it aims to maximise the utilisation of limited radio resources. It can be very complex in wireless networks as users are moving and their traffic pattern changes rapidly. Therefore, the critical dimension in a radio network is the allocation and use of the bandwidth in the radio cells in order to avoid local congestion or degradation of wireless link.

The work in [91, 92] used intelligent agents to control first generation mobile networks resulted with a distributed resource allocation scheme. It offered an efficient solution for resource allocation under moderate and heavy loads. The work in IST Project SHUFFLE [59], and the concept is extended to 3G networks. Later in [15-17], the author uses distributed intelligent agents and smart antenna technologies to optimise radio resources for WCDMA networks. So far, the previous works has been restricted to cellular based networks. The first approach uses intelligent agents and multi-agent systems to control and optimise the radio resource for overlapping WLAN systems.

4.3 Agent platform

Although a lot of agent-related research has taken place during the past couple of decades [93], the definition of the most appropriate architecture for multi-agent systems varies. At present, the most prominent [94] variants seem to be:

- Common Object Request Broker Architecture (CORBA) from The Object Management Group (OMG) [95];
- Knowledge-able Agent-oriented System (KAoS) [96];
- General Magic (a commercial endeavour researching mobile agent technology) for electronic commerce [97]; and
- The IEEE Foundation for Intelligent Physical Agents (FIPA) [98].
FIPA is now the eleventh Standards Committee of the IEEE Computer Society, but previously was in independent foundation promoting software standards for agent technology. The FIPA architecture was chosen for this research because it was publicly available. It [98, 99] specifies the interfaces of the different components in the environment with which an agent can interact, i.e. humans, other agents, non-agent software and the physical world. A particular feature of FIPA is the standardization of the agent communication, that is, the specifications for the “language” that enables agents to communicate with each other.

Agents are often implemented on an “Agent Platform” that offers a basic set of services: the Agent Management System (AMS), the Directory Facilitator (DF) and the Agent Communication Channel (ACC).

Agents must register to the AMS of an Agent Platform; they can also register to the DF, which enables them to offer their services to other agents. However, the DF registration is optional. Agents communicate using the Agent Communication Language (ACL) over the ACC between agents on a platform and between platforms. Figure 4.2 shows the reference model for the management of the FIPA agents.

Figure 4.2: FIPA agent management reference model\(^8\).

The AMS is the core of an Agent Platform and is responsible for registering/deregistering agents by calling the AMS message method with a registration request. The DF offers services similar to those of the AMS and acts as a “yellow pages” directory, where agents willing to offer their services in a dynamic manner to other agents may register.

The FIPA defined communication model [98] is presented in Figure 4.3: agent communication is accomplished through the use of ACL (which is a two layered structure composed by content language/ontology and the communication language.

![Figure 4.3: FIPA compliant Communication model](image)

Ontologies are schemes for describing concepts and their relationships [101]. Once interacting agents have a common ontology, they will use this ontology to interpret communications interactions, and in this way they will each be able to understand the actions of the others. The content language is used to combine terms in the ontology into sentences that mean something to the agents. The ACL is a language with precisely defined syntax and semantics and pragmatics that carries the sentences of the content language between agents. The ACL defines the types of messages exchanged, enabling the agents to accomplish complex tasks, such as negotiation and auction.

For this research the JADE [102] software framework for Intelligent Agents was used as it is a popular implementation of FIPA. JADE is implemented in Java and simplifies the development of a MAS through middleware that complies with FIPA specifications and through a set of tools that supports debugging and deployment. The agent platform can be distributed across different machines, not necessarily using the same OS, and the configuration can be controlled via a remote graphical user interface. The configuration can be changed at run-time by moving agents from one machine to another one, as and when required.

The communication architecture offers flexible and efficient messaging, where JADE
creates and manages a queue of incoming ACL messages, private to each agent; agents can access their queue through a combination of several modes: blocking, polling, timeout and pattern matching. The full FIPA communication model is implemented.

JADE features:

- A FIPA-compliant Agent Platform, which includes the automatic activation of AMS, DF and ACC at agent platform start-up.
- A distributed agent platform that can be split over several hosts. Only one Java application, and therefore only one Java Virtual Machine, is executed on each host.
- Different types of host computer: the host can be a PC or a mobile, embedded system like a PDA, mobile phone or the control unit of a base station. Agents are implemented as one Java thread and Java events are used for lightweight communication between agents on the same host. Parallel tasks can be still executed by one agent, in a more efficient way than the Java Virtual Machine does for threads.
- A defined programming interface to the DF.
- A transport mechanism and interface to send and receive messages.
- FIPA-compliant Internet Inter-Orb Protocol (IIOP) protocol to connect different agent platforms;
- Lightweight transport of ACL messages inside the same agent platform as messages are transferred encoded as Java objects, rather than strings. If messages are to go between platforms the message is automatically converted to the FIPA compliant string format, so that implementers only need to deal with the same class of Java object.
- A library of FIPA interaction protocols.
- Automatic registration of agents with the AMS
- A GUI to manage several agents and agent platforms from the same agent.

The security model of the JADE platform [103] allows implementers to manage authentication of users and agents through authentication and authorization procedures. The secure communication between the components is based on the Secure Socket Layer (SSL) protocol [104], which is a general-purpose protocol for ensuring the authentication and encryption of TCP connections.
The flexibility of the JADE platform allows different Access Point Agents to run on separate hosts or on one standalone workstation.

The agent container is a multithreaded execution environment consisting of one thread for every agent plus system threads spawned by the RMI runtime system for message dispatching. A special container acts as the front-end, running management agents and representing the whole platform to the outside world. A complete Agent Platform is then composed of several agent containers as shown in Figure 4.4. Distribution of containers across a computer network is allowed, provided that RMI communication between their hosts is preserved.

![Figure 4.4: Software architecture of the JADE Platform](image)

**4.4 Agent-based negotiation framework.**

**4.4.1 Architecture**

The combination of Intelligent Agent and Multi-Agent Systems (MAS) has been widely used in cooperative environments like e-commerce [105], distributed Artificial Intelligence (AI) [106], new IP routing protocols [33] and resource management of CDMA networks [59]. The advantage of using MAS is that it provides a distributed, robust platform that allows agents to act both individually and cooperatively. Within the WLAN simulation, these agents are AP agents and antenna agents; they all have their own initiative and co-operation ability as explained later.

AP agents communicate with each other through the MAS platform and change the radio coverage patterns co-operatively through the Antenna Agents depending to the geographical load distribution. A Simulation Agent initialises the simulation...
environment, provides the program interface, and collects the resulting data; it also acts as a directory facilitator to allow AP Agents to find their neighbours (Figure 4.5)

**Figure 4.5: The Multiple Agent System for WLAN resource management**

The key part of the whole system is the AP Agent and its associated Antenna Agent as shown in Figure 4.6. The AP agent performs the negotiations to satisfy the traffic changes and also the mutually agreed patterns (with neighbouring AP agents); the Antenna Agent performs the power and phases changes for the four antenna sectors, thus producing coverage generation.

**Figure 4.6: Access Point Agent and Antenna Agent basic structure.**

There are nine types of message involved in the negotiation and commitment processes:
**AFH (Ask For Help):** this message initiates the coverage negotiation process and contains the area for which an AP agent needs help from neighbouring AP agents.

**RAFH (Reply AFH):** message returned from adjacent AP agent, with the “price” to cover the area in the previous AFH message; this message is used to end a coverage negotiation.

**RFC (Request For Commitment):** message used to start the commitment process after negotiations reach an agreement.

**AOC (Acknowledgement of Commitment):** message answering the RFC message if the required hypothesis is still valid and the AP agent is not involved in any other transaction.

**CYP (Change Your Pattern):** when the initial AP agent receives all the AOC messages, it sends CYC message to the relevant base station agent to change coverage.

**CMP (Change My Pattern):** when an AP agent receives a CYP message, it sends a CMP to its own associated Antenna Agent to change the coverage pattern physically; the negotiation initiator will also send this message to its own Antenna Agent to change coverage pattern

**Scheme Expires (SE):** if the hypothesis has expired or the base station agent is currently involved in another transaction, a SE message is sent back to the RFC requester.

**Cancel Transaction (CT):** message used to cancel an ongoing transaction.

**Transaction Cancelled (TC):** message used to acknowledge the cancellation of the specific transaction.

### 4.4.2 Example negotiation

The concept of the negotiation scheme is derived from the work of Du [19]; however, in Du’s work, the negotiation scheme is designed for use with WCDMA networks. However in WLANs, the fundamental service provision mechanism is different and the negotiation scheme has been significantly adapted in order to work in a WLAN and multi-path environment.
Each AP agent monitors its own traffic load in real time; when the overall traffic load is over a certain threshold (95% of the maximum data rate supported by the AP is used here as it is regarded as a heavy load for an AP), a negotiation procedure is triggered.

A simple illustrative example of this process is shown in Figure 4.7. When the AP agent (AP3) detects its utilisation exceeds the threshold T, it initialises a pattern negotiation. The local optimiser in AP agent first proposes a set of local coverage hypothesis and sorts them according to some evaluation function (explained later in section 4.4.3).

In this example, 2 hypotheses (schemes) are included for easy understanding: scheme 1 and scheme 2. The negotiator creates AFH messages according to the coverage losses in each of the coverage hypothesis (these hypotheses also contain the STA information in those lost areas), and sends these AFH messages to the relevant adjacent AP agents. Suppose AP3 has some coverage losses in the area near AP1 and AP4; AP3 sends two AFH messages to both AP1 and AP4, i.e. AFH(1) and AFH(3). These two receivers have to propose some local coverage hypotheses that cover the losses of AP3 and if they have enough free resource (including free capacity and transmitting power), RAFH messages will be returned immediately, like RAFH(1) and RAFH(2). However, if their free resources are not enough to help AP3, this process will involve further negotiation with other AP agents, like the AFH(3) from AP4 to AP5. After AP4 gets the required RAFH(3), it returns AP3 RAFH(5) with accumulative calculated price.

Once the AP3 gets enough RAFH replies from the other AP agents involved, its executer (the component responsible for commitment process) chooses the one with the cheapest accumulative price (in this example hypothesis 2, as the scheme 1 involves further chain negotiations between AP4 and AP5 thus increasing the price), and starts the commitment by sending RFC messages to all the AP agents involved in this negotiation to commit their promised hypotheses. These AP agents, like AP1 and AP4, check the validity of the schemes and reply to the original AP agent AP3 with the confirmation message, AOC. If AP3 receives all the AOC replies, it sends CYP messages to all the AP agents involved. Finally, AP agents send CMP messages to their own antenna agents, the antenna agents then control signals to the electronic attenuators of the affected sectors in order to change coverage patterns.

After commitment, all messages relative to a negotiation are discarded, and the states for involved agents are reset to “idle”. They are then ready for further negotiation and
commitments.

Figure 4.7: The basic negotiation sequence diagram

As seen in this example, whenever an AP agent observes its utilisation exceeding the threshold T, it has to lose some coverage and STAs to reduce its utilisation, and
negotiations between adjacent base station agents are performed. If the helping AP agents in the immediate neighbourhood do not have enough free resource (capacity or sector power) the coverage negotiation may spawn further negotiations. Message loops are checked to eliminate potential deadlocks and hop counters are used to avoid chained coverage negotiation going to fare. The current hop limit is two. Since radio power decreases very quickly with distance, AP agents managing cells that are far apart are reasonably independent. This is also a factor that helps to achieve timely responses for negotiations, so that most of the negotiation tasks can reach good agreements quite quickly.

4.4.3 Negotiation process

Unlike cellular systems, in 802.11 the “handover” is initialised only by STAs. In the 802.11 MAC standards, the station periodically scans for all available radio channels and APs and then tries to associate/re-associate to the AP that provides the best signal strength and quality [39]. When sensing a stronger signal, it first disassociates from the currently associated AP and then initiates the association process to the “better” AP with stronger signal quality, thus achieving “handover”. Most WLAN vendors’ products already support seamless hand-over so that the application layer connection will stay connected during the whole handover process [37].

It is also worth mentioning that in order to reduce the number and frequency of STA re-associations to other APs, many 801.11 WLAN adaptors implement a mechanism called “stickiness”, which means that the client card will stick to the original associated AP even it has found a better signal quality AP. It will only re-associate to the better AP when the signal quality of the current one drops to a certain level or the new AP shows a very strong signal level. This is an issue for the co-operation system as the basic principle assumes the STA will always associate to the best AP it detected. A mechanism to overcome this issue is to use the AP’s de-associate process: that is the AP can actively de-associate the target STAs by sending them de-associate frames; those STAs will then re-scan the available APs and select the best AP (with best signal strength and quality) as desired.

As the coverage is no longer fixed, it is inadequate to model the area of AP coverage simply by a hexagonal cell. In order to realise the negotiation, some sort of model has to be used for coverage patterns. The maximum coverage (frontier) is the outreach of
the antenna patterns when all sectors are set to full power. Any demand outside that frontier is of no direct interest to the AP, as it cannot service it. Within the frontier the area is divided into locations using polar coordinates, as shown in Figure 4.8. Each location (the smallest division), named as Quantisation Cell (QC), can contain many associated STAs (the QC concept was first used in Du’s work [19], in his work, each QC can contain many CDMA terminals).

Before the start of negotiation, the local optimiser of the heavy loaded AP proposes several hypotheses that could reduce the AP’s load under the threshold T. In Figure 4.8, two example hypotheses are shown: hypothesis A involves dropping power in two sectors and consists of three QCs losses and hypothesis B involves one sector and one QC loss.

The local optimiser of AP agent proposes hypotheses based on reducing the power in one or more sectors. As a result, some QCs will have received signal power lower than the minimum service power. The STAs in these QCs will lose their connection to the original AP and re-scan for better signal AP; if found, the STAs will associate to the new AP and the intended hand over is achieved.

The AP agent will propose several hypotheses and assess them according to “costs”. The cost of the hypothesis is defined by how much disturbance the hypothesis will
bring to the current AP: the lower the power change and the smaller the number of STA handovers means the smaller is the disturbance to the current AP.

It is defined by (4.1):

\[ P_i(A, j) = w_1 \cdot PowerChange + w_2 \cdot STALosses \]  

(4.1)

Where \( P_i(A, j) \) represents the cost for a hypothesis (Hypothesis A) to \( AP_j \), \( PowerChange \) (in dB) is the total power decrease proposed by Hypothesis A, \( STALosses \) is the total number of STAs that will be handed over to neighbouring APs in Hypothesis A. \( w_1, w_2 \) are weights. The weights are chosen to balance the effectiveness of power decrease and STA losses to the cost of a hypothesis. \( PowerChange \) and \( STALosses \) can be expressed as (4.2)

\[
\begin{aligned}
\text{PowerChanges} &= \sum_{i=0}^{i<4(\text{all sectors})} PowerChange_i \\
\text{STALosses} &= \sum_{i=0}^{i<4(\text{all sectors})} \sum_{k=0}^{\text{all QC losses in sector}} N(k, j)
\end{aligned}
\]  

(4.2)

Where \( PowerChange_i \) is the ith sector's power change proposed by the hypothesis. \( N(k, j) \) is the number of STAs being served \( AP_j \) by in \( QC_i \). The AP should keep both \( PowerChange \) and \( STALosses \) to the minimum in order to avoid too much disturbance to the system. This is achieved by processing the hypotheses with the lowest local cost first.

The local optimiser keeps proposing new coverage loss hypotheses up to an upper limit (currently four as explained before). Each hypothesis consist of 1 or more QCs that could relieve the AP’s traffic under threshold \( T \).

\[ Load_j - Load_{\text{hypothesis}} < T_j \]  

(4.3)

The evaluator checks the coverage losses in the hypotheses, and the negotiator creates
up to six\(^9\) AFH messages for possible losses and sends them to adjacent AP agents. Each AFH message may contain one or more coverage loss hypotheses. To simplify the system, only the closest AP agent is selected as the helper for each request. As one hypothesis’s coverage losses may need more than one neighbouring AP to help, the hypothesis may be sent to more than one helper AP - in the example of Figure 4.8, hypothesis A needs two helper APs to cover the lost areas. The process of coverage negotiation is shown as Figure 4.9.

---

\(^9\) There are six immediate neighbours for each access point, this is the common topology of large deployment of WLANs, as shown in Figure 4.7.
When the helper receives AFH messages, its local optimiser also proposes some local coverage hypotheses that arbitrarily cover the QCs stored in the AFH messages. If there is enough capacity and radiation power (towards those QCs) at the helper, RAFH messages with local costs will be returned back to the requester to finish this coverage negotiation. However, if there have to be coverage losses, further negotiations with the second level helper (except the requester) will be initiated, and the costs are accumulatively calculated. Any loops in helper requests have to be detected in order to eliminate potential deadlocks. Currently a maximum hop limit of two is implemented to prevent chained coverage negotiations from going too far. The local cost of Hypothesis B of helper $A_P^k$ is defined as (4.4).

$$P_2(B,k) = w_j \cdot \text{PowerChange} + w_4 \cdot \text{STALosses} \quad (4.4)$$

Where $P_2(B,k)$ is the local cost for a hypothesis (Hypothesis B) of helper $A_P^k$ to $A_P^j$. Again, $\text{PowerChange}$ and $\text{STALosses}$ can be calculated as (4.2). If the helper $A_P^k$ does need not any further help (i.e. no coverage losses need to be covered by further helpers), the second part of (4.4) will be zero and the cost will only be related to power changes at $A_P^k$. However, if it needs further help, it has to wait for the RAFH replies from chained helpers. After that, it calculates the costs accumulatively and returns the total cost to the initiator.

Finally, when the initiator receives enough RAFH replies, the negotiator picks up the cheapest solution according to the total cost (local cost plus helpers’ costs, as shown in (4.5)), and passes it to the executor for commitment.

$$P_{total} = P_1(A,j) + \sum_{k=1}^{\text{all helpers}} P_2(B,k) \quad (4.5)$$

### 4.4.4 Commitment process

In a distributed environment, a two-phase commitment protocol [107] has been widely used to synchronise the commitments for distributed transactions. A modified version of the two-phase commitment is used as the protocol for committing the agreements in this work, as shown in Figure 4.10.
Figure 4.10: State transitions of commitment

When the negotiation initiator finds a solution, it starts a distributed transaction to
activate this solution. The AP agent first checks which of the adjacent access point agents are involved with this transaction. It then produces and sends RFC messages to each of the helpers, which contain the ID of the helper’s hypotheses. It also sets itself as “in transaction” state.

After receiving the RFC message, if the helper is not currently in other transaction and the hypothesis is still valid, the agent returns an AOC message and set itself as “in transaction” too. The RFC and AOC messages may spawn if the helper needs further help.

When the initiator gets all the AOC replies, it creates and sends CYP messages to involved helpers asking them to finally change their coverage. Similarly, further CYP messages may be sent to second level helpers. When the helpers receive CYP messages, they send CMP messages to their associated antenna agent. The antenna agent will then start to change the relevant sector’s power and finally change the coverage.

However, if the RFC message arrives at a helper that is currently in another transaction, or if the hypothesis being asked for has expired, the whole transaction has to be cancelled. This helper sends SE and CT messages to the initiator and possible helpers respectively. Again, further SE or CT messages may be needed to cancel this distributed transaction. Those AP agents reset their states and acknowledge the cancellation with TC messages. After the initiator receives all TC messages, it starts another commitment process with the next best solution.

After all the agents commit their agreements, all the messages related to this negotiation are discarded, the agents set their states to idle and are ready for future negotiation and commitments.

4.5 Simulator for WLAN system with semi-smart antennas

4.5.1 Introduction

The negotiation algorithm is customised so that it can be applied to an 802.11 WLAN network with semi-smart antennas. In order to avoid producing unrealisable coverage, an access point’s coverage is divided into many small locations using polar coordinates. Each AP maintains its own coverage by the polar coordinates and negotiations with other APs are based on the QC as explained earlier.
In order to evaluate the performance of the negotiation algorithm in a WLAN network with semi-smart antennas, a good WLAN network simulator is needed. Most popular network simulation packages are not suitable for this work, as none of them have considered both hot-spot traffic and semi-smart antennas; also, most simulators are frame-based and designed to simulate the detail of the MAC or PHY protocol. This makes it very difficult to simulate a large scale multiple-AP network with many mobile users and hot spots like the case here.

In order to test the agent-based negotiation for WLAN system, a comprehensive WLAN network simulator including some advanced features, such as supporting indoor radio propagation, configuration file, individual power control of semi-smart antenna, various traffic types (browsing, email, file download) and user mobility was built. The simulator is capable of simulating WLAN networks with unevenly distributed traffic (hot spots), as this is the main reason of investigating the negotiation algorithm. Validation of the simulator was performed to assure the correctness of the simulation results (in Section 4.5.3).

4.5.2 System architecture

The simulator contains three main parts: data input, results output and simulation core. Consider each part in turn:

Data input

1. Reads radio propagation file as the base of the indoor propagation model.

The simulation has the ability to read simulated/actual radio propagation readings for specific locations (such as campus, airport lounge etc) and is able to initialise the QCs’ signal power readings accordingly. This interface allows the simulation to be run for a specialised geographical setup, how this is done is discussed in section 4.5.4.

2. Read semi-smart antenna configuration file

The current semi-smart antenna is a four-sectored individual controlled patch antenna. The first simulation was using ideal radio coverage pattern for each sector (each sector has a 90 degree coverage angel). Later simulations will be able to read the sector antenna’s propagation file and also not be limited to four sectors.
3. Read general simulation configuration file.

This includes other setup steps such as location of hot spot, location of access points, number of STAs and other simulation parameters etc.

**Results output**

1. Write traffic load on each simulation time advance.

Each access point’s current traffic demand and load were written to a result file for evaluation of the negotiation algorithm.

**Simulation core**

The key part is the simulation core; it performs calculations for traffic demand, user mobility and cooperates with agent negotiation system. The flow chart is shown in Figure 4.11.
The simulation is performed iteratively to calculate the system capacity for the traffic configuration in each time epoch. In the scenarios considered here, mobile WLAN users come into the simulation area and form a traffic hot spot. The current hot spot is generated by producing a mobile user’s X and Y coordinates by a normal distribution with the mean value at the hot-spot centre. A more realistic traffic hot spot can be generated by reading from user location files. All users in the network can only be

Figure 4.11: flow chart of the WLAN network simulator.
associated to one access point at a given time. This constraint is achieved in the simulation by comparing the perceived signal strengths for all access points. The mobile user will associate to the access point with the best signal quality. Upon association, the user will be able to communicate with the AP directly to access Internet resources. The data rate for the mobile user is defined by the technology used at the access point (802.11b or g).

The whole bandwidth of an 802.11 access point is shared by all associated mobile users. For an individual access point, if the demand does not exceed the maximum data rate (net data rate or throughput), each user will have an achieved data rate that is equal to the demand. However, if the demand exceeds the maximum for the access point, all user will have a lower achieved data rate than traffic demand and an individual user will have an achieved data rate that is proportional to its traffic demand.

When the current simulation time equals the minimum negotiation time interval, all access points will check their current traffic load and perform the negotiation if necessary. If the negotiation results change the antennas’ coverage, the simulator will update the radio coverage accordingly. Some mobile users will then be re-associated to other access points due to the change of radio environment.

As the simulator is written in a multi-thread and multi-agent environment, each access points is represented as an individual software agent and the whole simulation is easy to deploy into a distributed environment.

4.5.3 Simulation validation

To demonstrate that the simulator is performing correctly and the simulation results are trustworthy, validations were carried out during and after the development of the simulator. These validations are performed on the three main actors of the simulator: the STA, the AP and the whole network.
1. **Validation rule**: a STA is served by only one AP at any given time.

   **Description**: In 802.11 WLAN, a STA can be served by only one AP, the STA is not allowed to associated with more than 2 APs and communicate with them. This is not the same as in CDMA network, in where soft hand-over allows a mobile station to be served by more than one base station.

   **Module**: STA.

   **Conditions**: during the association control process and re-association control after negotiations.

   **Result**: Pass

2. **Validation rule**: a STA is always associated to the AP that supplies the best signal quality, whenever it starts an association or re-association process.

   **Description**: Although there is no specific rules in 802.11 to define the criteria for selecting an AP during the association and re-association processes, signal-level based selection is usually deployed (e.g. by RSSI or SNR) in current 802.11 devices [37]. This research uses this approach as well.

   **Module**: STA.

   **Conditions**: during the association control process and re-association control after negotiations.

   **Result**: Pass

3. **Validation rule**: a STA’s achievable data rate is equal to or lower than its link data rate.

   **Description**: 802.11 devices can communicate at different speed depending on signal quality, so for an individual STA, the maximum achievable data should not exceed the link data rate.

   **Module**: STA.

   **Conditions**: during the association control process and re-association control after negotiations, also during calculation of throughput of STA and AP

   **Result**: Pass
4. **Validation rule**: the SNR value of any served STA (by any AP) should be larger than or equal to the threshold.

**Description**: in order to serve a STA at a certain data rate, the STA must maintain a certain level of SNR or RSSI. Below the threshold, the STA will roll back to a slower data rate or even drop the connection.

**Module**: STA.

**Conditions**: during the association control process and re-association control after negotiations, also during calculation of throughput of STA and AP, the SNR is calculated to determine the link data rate of the STA.

**Result**: Pass

5. **Validation rule**: the transmission power at any sector should be in the adjustable range of power control.

**Description**: Currently the transmission power at any sector of an AP is limited to 0-20 dBm (1-100 mW). During the simulation, any adjustment of the antenna sector’s power should limited to the range

**Module**: AP.

**Conditions**: initially each sector of AP is set to 10 dBm; any adjustment of sector power is checked after any decision of changing antenna power is made.

**Result**: Pass

6. **Validation rule**: the load of each AP should equal to the sum of achieved data rate of all served STAs.

**Description**: in infrastructure mode, any communication is through the AP so the total load of AP is equal to the sum of data rate of individual serving STAs.

**Module**: AP.

**Conditions**: after each association, re-association process.

**Result**: Pass

7. **Validation rules**: if the sum of all serving STAs’ demand is greater than the APs maximum data rate then the achieved data rate for each STA is less than its
Description: an 802.11 AP does not have a “Call Blocking Rate” concept as it will take any STA that request association, so if overloaded, every serving STA (by that AP) is affected and will get a reduced data rate.

Module: AP.

Conditions: after each association, re-association process.

Result: Pass

All the validations have been shown to be true for ten simulation runs for both conventional and negotiation based networks.

4.5.4 Measurement and pass loss calculation hybrid method

Most WLAN systems are deployed in an indoor environment, so the free space radio exponential path loss model fails to represent the actual radio propagation characteristics because of severe multi-path effects. Experimental work has been published [108] showing that signal strengths can dramatically change for a distance of as little as a single wavelength. Even worse, the radio propagation is site-specific (affected by layout of building structure, building material etc). For this research, the calculation of the signal strength between sample-points utilises the indoor log-distance path loss model described in [109] (Equation (4.6)) combined with measurements.

\[ PL(dB) = PL(d_0) + 10n \log\left(\frac{d}{d_0}\right) + X_\sigma \] (4.6)

Here the value of \( n \) depends on the frequency used and the surroundings and building type, and \( X_\sigma \) represents a normal random variable in dB having a standard deviation of \( \sigma \) dB.

With a 802.11b WLAN that uses the 2.4 GHz ISM band and a typical indoor building, the path loss propagation model (4.6) can be written as (4.7) [110]

\[ PL(dB) = 40 + 35\log D + X_\sigma \] (4.7)
Here D is the distance between the transmitter and receiver in metres.

To summarise the approach of modelling used in this thesis, the indoor propagation model has two steps:

- Selecting a series of sample points and measuring the signal strength using a WLAN client card (it actually records the Radio Signal Strength Indication (RSSI) parameter available in the PHY layer).
- The rest of the area is calculated from these samples by using the indoor propagation model (4.7).

As noted earlier, the location of the WLAN clients can be achieved by using the Ekahau WLAN positioning engine, which uses a patented probability method to estimate the user’s location according to the signal strength received at a normal WLAN card. The overall accuracy can be up to few metres and is acceptable for positioning purposes. An alternative method would be to use GPS, but this requires extra equipment and has limitations indoors.

### 4.6 The simulation for the agent-based approach

Based on the system described, an agent-based WLAN radio resource management simulation system was developed. Client stations are distributed across the whole area and a hot spot (congestion area) is generated to test the efficiency of the negotiation process. The simulation has the following assumptions:

- Each AP is mounted with the proposed 4-sector semi-smart antenna, each sector having individual power control in order to produce a flexible coverage pattern.
- 16 APs are laid out on a 4×4 hexagonal pattern (similar to a cellular system), so one AP can have up to six neighbours.
- The maximum data rate for each AP is set to 11 Mbit/s and a maximum net throughput of 6.8 Mbit/s for 802.11b and 24.7 Mbit/s for 802.11g.
- 96 client stations or STAs are normally distributed to produce a hot spot centred at AP 5.
- Each STA has static or dynamic data demand; the simulation results for the two settings are evaluated individually.
Radio propagation characteristics and coverage area for each AP comes from the setup file; it imitates the actual system where a site survey phase with sampling would produce the data.

As explained earlier, the client stations are presumed to always associate to the AP with the strongest signal strength.

![AP Layout of the simulation.](image)

The whole simulation system can be easily adapted to multiple machines where each machine represents an individual AP agent. The messages used in the system are based on FIPA ACL message [98]. With addition of Ontology content language support, the negotiation protocols are scaleable and easy to use.

### 4.6.1 The Result

In order to examine the system performance, an evaluation parameter called “load/demand” is used. The load/demand is the achieved throughput divided by the demand (application data rate demand). In the result here it is aggregated over all users. The demand is defined as the user’s application data rate without the wireless link, i.e. the STA is connected to the Internet directly. The user’s load is defined as the user’s application’s achieved throughput when using WLAN connection. Two results are shown in Figure 4.13 (for fixed demand from each user) and in Figure 4.14 (for random demand) show that with AP negotiation and dynamic changing of radio coverage patterns according to the traffic distribution, a significant increase in the overall WLAN system performance is seen.
The results in Figure 4.14 were generated with random data demand: each station’s demand is uniformly distributed (between 0 to 2.0 Mbit/s) to simulate multiple users having different data rate applications. For the static demand simulation, an average data rate of 1.1 Mbit/s was set for all user stations (this matches the average data rate measured in section 5.3).

Both scenarios show that with the load balancing, the data traffic in congested “hot spot” areas can be balanced effectively.

The radio coverage patterns of the APs around the “hot spot” can be seen in Figure 4.15 – this is after the result of negotiation for the random data scenario. The results shows how the overloaded AP shrinks its own coverage pattern and the other APs enlarge theirs to help.

Figure 4.16 shows the results of simulated throughput performance improvements both for 802.11b and 802.11g. From the figure it can be seen that with negotiation and dynamic changing radio patterns, the 802.11b WLAN can perform almost as well as 802.11g WLAN in its normal configuration (with omni-directional antennas).

Figure 4.13: Simulation result of using static user demand.
Comparation between using and Not using agent-based dynamic negotiation radio pattern control (random user data demand 0 - 2 Mbit/s)

Figure 4.14: Simulation result of using variable user demand

Figure 4.15: Screen shot of part of the simulation GUI
Both sets show that as the number of users (and hence the load) increases, the data rate (expressed here as load/demand) available to users decreases. However, with dynamic agent-based radio pattern control, the congested APs negotiate and co-operatively change their radio propagation patterns. The performance is very much better than the system without dynamic radio control.

These results are from 10 independent simulations with different seeds for generating the random numbers, 10 simulations being used to eliminate the effects of different seeds. Both results show the limits of results produced from the different random seeds as confidence bars, with the curves being the average across all results. However, Figure 4.14 additionally shows the result from a single set of results - this demonstrates that the improvement for the average and for this single result is substantially the same.

Figure 4.17 shows the results expressed in terms of improvement (improvement of the
The system can handle approximately double the number of users at full capacity, i.e. with the carried load being the same as the demand.

With greater congestion the system cannot carry all the demand, but the improvement continues to increase.

As the congestion grows the benefit will naturally start to tail off, since there is only a finite capacity available and even better distribution of that capacity will not be able to cope with unlimited demand.

Figure 4.17: Improvement obtained with negotiation (802.11b)
This chapter proposed a novel agent based negotiation algorithm that can effectively balance traffic between multiple APs when there are hotspots occurring in the system. Agents representing APs negotiate with each other and reach a mutual agreement on how to change the antenna radiation patterns. Some mobile users in the heavily loaded APs will drop their connection with those APs and re-associate themselves to neighbouring, more lightly loaded APs. This shifts load away from the congested APs.

A traditional 802.11 network can be easily upgraded to an agent-negotiation-controlled 802.11 system through a firmware and software upgrade (embedding agents into firmware and driver software) and replacing the omni-directional antenna with (fairly cheap) power-controllable semi-smart antennas. Initial simulation results show such a system can improve the system performance (in terms of overall system load/demand) by up to 40%.
Chapter 5  Simulation under constrained environment

5.1 Introduction

In this chapter, research is described to take into account a more accurate path loss model, a multiple-user traffic model and a interference model, as well as using a more realistic user scenario with a constrained indoor environment.

As this research is concerned with deploying a WLAN system over a large area and optimising network throughput in congested scenarios, the impact on application level throughput for other users in the same cell needs to be investigated, as well as interference from other access points and users in other cells. The WLAN multi-access uses DCF (CSMA/CA) mechanism which is very different from traditional cellular systems (such as GSM or CDMA) and needs to be investigated.

The previous simulation was based on a very simple throughput calculation algorithm, which assumes all users in a WLAN Basic Service Set (BSS) share the total throughput of the WLAN fairly; this assumption is good for getting a general view of the network situation. However for more realistic, accurate results, the simulator needs to be improved by refining the user data rate model.

Most research papers on WLAN are on theoretical analysis and data frame level simulation of WLAN’s MAC and PHY algorithms [111-114]; very few papers have actually discussed the deployment and design of public large scale WLANs, especially with many mobile users [115].

To provide a more accurate representation, the simulator was enhanced by adding:

- a more accurate user data rate calculation in a multi-access Public WLAN (PWLAN) environment;
- the ability to take into account the interference caused by neighbouring APs and users; and
- configuration files to define a floor map and a more complicated user scenario (such as hot spot production).
5.2 Data rates achieved in a multi-user environment

As defined in the IEEE 802.11 standard, a station (STA or AP) needs to decide the data rate of a ready-to-send frame. The most common method is to measure the SNR of the most recent received frame to decide the data rate to be sent. The new data rate calculation (for simulation) for each user is based on an approach by [115-118] which is derived from empirical data, using a formula to map SNR to application throughput based on numerous experiments.

There are two models – piecewise linear and exponential [116]:

Piecewise linear:

\[ T = \begin{cases} T_{\text{max}}, & \text{if } \text{SNR} > \text{SNR}_c \\ A_p \cdot (\text{SNR} - \text{SNR}_0), & \text{if } \text{SNR} \leq \text{SNR}_c \end{cases} \]  

(5.1)

with

\[ \text{SNR}_c = \frac{T_{\text{max}}}{A_p} + \text{SNR}_0 \]  

(5.2)

Exponential model:

\[ T = T_{\text{max}} \left(1 - e^{-A_e(\text{SNR}-\text{SNR}_0)}\right) \]  

(5.3)

\(T\) is throughput. \(T_{\text{max}}, \text{SNR}_0, \text{SNR}_c\) and \(A_p / A_e\) are constants that are vendor and application specific. \(T_{\text{max}}\) is the throughput saturation level which results from the SNR going beyond the critical threshold \(\text{SNR}_c\).

\(\text{SNR}_0\) is the SNR where throughput is zero. In the piecewise model of (5.1), \(A_p\) is the slope of the line when \(\text{SNR} \leq \text{SNR}_c\). In the exponential model of (5.3), \(A_e\) describes the rate at which the throughput reaches maximum.

Using parameters from recommended in [115], the simulator was modified to use the exponential model with the following values: \(T_{\text{max}} = 5.3; A_e = 0.069; \text{SNR}_0 = 5.4\)

By applying the above parameters to (5.3), an 802.1b data rate curve approximation can
be obtained as shown in Figure 5.1

![Figure 5.1: the exponential model of 802.11 b data rate approximation](image)

The model described above can be extended to a multi-user environment model by subtracting throughput loss factors due to the interaction between multiple users. When \( N \) users are present, the throughput prediction for user \( i \) is denoted as “\( T_{N\text{-user, } i} \)”, which can be modelled as [115]:

\[
T_{N\text{-user, } i} = T_{1\text{-user}}(SNR_i) - \sum_{\{k|kN,k \neq i\}} [T_{loss\text{-compete}}(SNR_k) + T_{loss\text{-hidden}}(C_i, C_k)] - T_{offset} \tag{5.4}
\]

\( T_{1\text{-user}} \) is the throughput of an individual user without the presence of other users, as presented in (5.1) and - the single-user empirical model. \( SNR_i \) is the SNR of the user under consideration, i.e., the \( i \)th user. \( SNR_k \)'s are the SNRs of other users that are sharing a common AP with the \( i \)th user. \( T_{loss\text{-compete}}(SNR_k) \) is the throughput loss due to competition between the considered \( i \)th client and the \( k \)th user for access to the AP. Equation (5.4) describes the fact that the more the \( k \)th user is experiencing poor...
throughput, the more likely it is that the user will interfere with, and decrease the throughput of the $i$th user.

$$T_{\text{loss-compete}}(\text{SNR}_i) = A_{\text{compete}} \times [T_{\text{max}} - T_{\text{user}}(\text{SNR}_i)]$$  \hspace{1cm} (5.5)$$

and

$$T_{\text{loss-hidden}}(C_i, C_k) = A_{\text{hidden}} \times PL(C_i, C_k)$$  \hspace{1cm} (5.6)$$

$T_{\text{max}}$ is the saturation throughput obtained from single-user measurements. $A_{\text{compete}}$ is a unitless constant that models the degree of throughput competition between the $i$th user and the other users. $T_{\text{loss-hidden}}$ is the throughput loss due to the so-called “hidden-terminal” problem [114, 119].

In the simulation here, the following parameters based on experiments [115] were used: $A_{\text{compete}} = 0.15$; $A_{\text{hidden}} = -0.72$ and $T_{\text{offset}} = 180$.

### 5.3 Modelling WLAN user traffic

This section introduces how user traffic is modelled, starting from considering the normal user traffic pattern on the Internet as the basis for modelling WLAN user demand. A series of tests of download various common files from different websites were conducted to provide some evidence of the characteristics.

At present, most applications over WLANs are non-real-time data traffic (rather than, interactive services like VoIP – although that may change in the future). The simulation assumes connections, which in WLAN applications map to a series of data sessions.

There is strong evidence that the session arrival process is Poisson. That is, human Internet users seem to operate independently at random when initiating access to certain Internet resources (e.g, Internet downloading). This observation has been noted for several network applications. For example, Paxon and Floyd [120] studied telnet traffic and found that the session arrival process is well-modelled with a Poisson process, though with a time-varying rate. Similarly, Arlitt and Williamson [121] found that the user requests for individual Web pages on a Web server can be modelled well by a Poisson process.

The user data traffic is modelled as an ON-OFF process [122]. It consists of 2 states,
active (ON) and silent (OFF) stage, with a transition rate $\mu$ from On to OFF and $\lambda$ from OFF to ON stage. Figure 5.2 illustrates the ON-OFF model.

![Figure 5.2: Traffic model for user demand](image)

Figure 5.2: Traffic model for user demand

In this work, it is assumed that the user data traffic has a variable demand during the ON period and the ON periods represent 45% of the time. Referring to the approach in [123] to simplify the simulation, an activity factor of 0.45 has been used. The activity factor is the ratio of the ON period over the total time. The gap time between user data sessions and the holding time each have a negative exponential distribution and the session arrivals follow a Poisson distribution. The mean holding time for a session assumed to be 180 seconds. All data traffic arrival distributions and call holding distributions are independent of each other.

As the demand of each data session is variable, it is necessary to justify the demand in the real world. A collection of various popular http and FTP downloads from servers all over the world have been used to justify the session download speed (assuming that most of the traffic in WLAN is down-link). Each of the test items was downloaded three times at different locations and the results are shown in Table 5.1:

<table>
<thead>
<tr>
<th>Item</th>
<th>Location 1 (kByte/s)</th>
<th>Location 2 (kByte/s)</th>
<th>Location 3 (kByte/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Windows® Defender (Beta 2) x86</td>
<td>185</td>
<td>110</td>
<td>230</td>
</tr>
<tr>
<td>2. Google Earth</td>
<td>85</td>
<td>100</td>
<td>160</td>
</tr>
<tr>
<td>3. Skype 2.5 beta</td>
<td>17</td>
<td>210</td>
<td>136</td>
</tr>
<tr>
<td>4. Cordis European website</td>
<td>170</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>5. FileZilla Ftp client: host North America</td>
<td>113</td>
<td>150</td>
<td>130</td>
</tr>
<tr>
<td>6. FileZilla Ftp client: host Dublin, Ireland</td>
<td>75</td>
<td>25</td>
<td>135</td>
</tr>
<tr>
<td>7. FileZilla Ftp client: Taiwan, China</td>
<td>220</td>
<td>180</td>
<td>55</td>
</tr>
<tr>
<td>8. FileZilla Ftp client: Ishikawa, Japan</td>
<td>117</td>
<td>105</td>
<td>52</td>
</tr>
<tr>
<td>9. FileZilla Ftp client: Curitiba, Brazil</td>
<td>110</td>
<td>110</td>
<td>70</td>
</tr>
<tr>
<td>10. FileZilla Ftp client: Sydney, Australia</td>
<td>80</td>
<td>65</td>
<td>59</td>
</tr>
</tbody>
</table>
Table 5.1: download speed of various popular HTTP and FTP traffic.

<table>
<thead>
<tr>
<th>Item</th>
<th>Location 1 (KByte/s)</th>
<th>Location 2 (KByte/s)</th>
<th>Location 3 (KByte/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.  Mini City Guide Macau (government site), Macau, China</td>
<td>150</td>
<td>130</td>
<td>58</td>
</tr>
<tr>
<td>12.  London underground map (UK gov site) U.K.</td>
<td>180</td>
<td>86</td>
<td>150</td>
</tr>
<tr>
<td>13.  MSN messenger 7.5: msn UK website</td>
<td>220</td>
<td>204</td>
<td>150</td>
</tr>
<tr>
<td>14.  Adobe Arcobat Reader 7.0.1 update</td>
<td>170</td>
<td>184</td>
<td>270</td>
</tr>
<tr>
<td>15.  ICQ 5.1</td>
<td>110</td>
<td>110</td>
<td>180</td>
</tr>
<tr>
<td>16.  QDB 2.5, software</td>
<td>135</td>
<td>115</td>
<td>&gt;1MB (local)</td>
</tr>
<tr>
<td>17.  JSLAB - Java Scheme Library</td>
<td>32</td>
<td>210</td>
<td>1MB (local)</td>
</tr>
<tr>
<td>18.  Pdf 995 from MPI college</td>
<td>10MB/s (local)</td>
<td>10mb (local)</td>
<td>59</td>
</tr>
<tr>
<td>19.  CUTE PDF from MPI college</td>
<td>10MB/s (local)</td>
<td>10mb (local)</td>
<td>59</td>
</tr>
<tr>
<td>20.  AIM from aol.com</td>
<td>110</td>
<td>170</td>
<td>n/a</td>
</tr>
<tr>
<td>21.  Ipswitch FTP site</td>
<td>56</td>
<td>67</td>
<td>215</td>
</tr>
<tr>
<td>22.  Lucasarts FTP site</td>
<td>160</td>
<td>185</td>
<td>120</td>
</tr>
<tr>
<td>23.  Sunsite FTP site</td>
<td>10</td>
<td>130</td>
<td>150</td>
</tr>
<tr>
<td>24.  Trumpet News Reader</td>
<td>23</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>25.  UK Winsock Archive</td>
<td>68</td>
<td>170</td>
<td>500</td>
</tr>
</tbody>
</table>

The download speeds of all items (except for local downloads) are counted and summarised as in Figure 5.3. From the figure it can be seen that the download speeds are distributed between 0 to 2.5 Mbit/s and the peak is between 1 to 1.5 Mbit/s. As the distribution of download speeds is similar to the probability density function of the beta distribution [124], the download speeds of different resources can be mathematically modelled as a beta distribution with $\alpha = 3$ and $\beta = 7$.

![Figure 5.3: Download speeds of various HTTP and FTP traffic session](image-url)
5.4 Modification and alternation to the negotiation algorithm

In the negotiation algorithm used earlier in the thesis, when an AP wants to hand over a client to a neighbouring AP, it must reduce the signal power to the client to be lower than the minimum service power. Then the client will initiate a new scanning process in order to find another AP that can offer the best signal power.

In a practical multi-AP environment, APs are usually placed much closer than the recommended layout and hence the negotiation needs to be modified to reflect the fact that there is virtually no “border” concept. The associated clients are much closer to the AP and as a result the AP will not always be able to reduce its radiation power and cause the target client to drop the link. Modification and improvement of the negotiation algorithm has been made.

To get round this problem, the negotiation is changed to be QC based instead of maintaining the “border” concept.

A simple diagram (Figure 5.4) shows the QC based negotiation concept. QCi belongs to AP1 and contains a mobile client. AP1 proposes to lose all clients in QCi and AP2 is willing to help. Initially the mobile client in QCi received signal power from AP1 P(AP1, QCi) and signal power from AP2 P(AP2, QCi). In order to hand over all clients in QCi to AP2, AP1 and AP2 will negotiate and mutually adjust radiation power towards QCi so that after adjustment, P(AP2, QCi) will be larger than P(AP1, QCi). In current 802.11 standards, although the mobile client in QCi will have sensed that AP2 has the strongest signal power, it will not always initiate the handover process from AP1 to AP2. However, AP1 can actively de-associate the client and force it to rescan for a better AP with stronger signal power.
In the new QC based negotiation process, when an AP’s traffic load is higher than the threshold $T$, it first proposes that some QCs are lost to reduce its burden. The selection of the QCs to lose is based on a utility function: the utility of a QC represents the serving importance for the AP. QCs with lower utility values have a higher chance to be selected as the ones to be removed.

Given that $QC_i$ is currently covered by access point $AP_j$, the probability of losing its coverage is proportional to how unimportant it is to the $AP_j$. If far away, $QC_i$ is less important to the $AP_j$ and hence is more likely to be selected as a removal candidate. If prospective helper $AP_k$ is lightly loaded, it has less difficulty to take over the STAs in $QC_i$, and $QC_i$ is more likely to be selected. More clearly, this can be formulated as the utility function $U(i, j)$

$$U(i, j) = \omega_0 \cdot I(i, j) + \omega_1 \cdot D(i, k)$$

(5.7)

Where $I(i, j)$ represents the importance of serving $QC_i$ for $AP_j$, and $D(i, k)$ represents the difficulty of the helper $AP_k$ to take over the STAs in $QC_i$. $\omega_0$ and $\omega_1$ are weights. $I(i, j)$ and $D(i, k)$ can be expressed as (5.8)

$$\begin{align*}
I(i, j) &= (1/\text{Distance})^2 \\
D(i, k) &= 1/(\text{Capacity}_k - \text{Load}_k + 0.001)
\end{align*}$$

(5.8)

Where Distance is the distance from $QC_i$ to $AP_j$, Capacity$_k$ is the capacity of the adjacent access point $k$ that is like to offer help. Load$_k$ is the current load of that
access point. To avoid dividing by zero, a small value is added in denominator.

The probability function (5.9) was used to represent the probability value of $QC_i$ to be selected as a removal by $AP_j$. It is the normalised form of $U(i,j)$ with exponential re-scaling of $\alpha$. The probability of $QC_i$ being selected as a removal candidate is inversely proportional to $U(i,j)$, hence proportional to $P(i,j)$.

$$P(i, j) = e^{\alpha(1.0-\frac{U(i,j)}{U_{\text{max}}})}$$

(5.9)

As the traffic does not change significantly between control cycles, it is not necessary to evaluate all locations within the coverage. Currently the AP agent only looks at 20% of QC's from the boundary.

If the selected QC(s) can reduce the traffic load of $AP_j$ under threshold $T$, a hypothesis is proposed consisting of the selected removal QC(s) and STA(s). The evaluation of the hypothesis and the negotiation process is similar to previous chapter, but instead of proposing reducing the QC(s) signal power to below the minimum service power, the current signal level at QC(s) is enclosed as part of the hypothesis. The helper AP agent will check the original signal power at these QC(s) and propose its own sector power change that could cause the signal powers at these QC(s) to exceed the old values. If the initiator AP agent approves the hypothesis, it will change the antenna's power and actively de-associate those STA(s) in those QC(s). The mobile STAs are forced to re-scan the environment and associate to the new AP with stronger signal power as expected.

This new step is to take into account the multi-path and overlapping environment in the coverage. The border concept used previously does not apply any more.

### 5.5 Multi-path radio propagation

The AP agent needs to know its own geographic radio coverage pattern as the basis for negotiation. It is necessary to know the detailed radio propagation map (radio propagation at every location of the simulation area) instead of using a simple path loss model for indoor multi-path rich environment. In practice, two methods are used to accomplish the radio propagation map: by using a site survey and by software
5.5.1 Using site-survey to generalise the radio propagation map

A site-survey\(^\text{10}\) based technique can be used to find the AP’s radio propagation in an indoor environment. This process can be shown in Figure 5.5.

![Diagram of site-survey process](image)

1. Measuring at sample points.
2. The generalised radio coverage pattern.
3. Apply the Polar coordinates.
4. The final representation of the radio coverage area in Agent knowledge.

**Figure 5.5: Polar coordinate representation of radio coverage pattern (only half circle of the radio coverage pattern to show the concept)**.

The steps in the process are:

- Select sample points distributed across the whole WLAN deployment; these points can be randomly selected geographically but more sample points and a more even distribution will increase the resolution and accuracy of the radio coverage pattern generated. An interval of 5-10 metres would be good practice.

---

\(^\text{10}\) The survey was performed using the Ekahau positioning engine: available at: http://www.ekahau.com. This is also used to track the location of individual STAs.
Use the 802.11 client card to record the signal strength at each sample point under maximum and minimum power control of all AP antennas. These raw data will be input to the agent system to help to calculate the radio patterns.

This process produces a generalised radio coverage pattern and by mapping this to 2-D polar coordinates, it is possible to represent simplified patterns in the agent system (refer to figure above). Using polar coordinates makes the complex radio coverage patterns mathematically easier to handle.

5.5.2 Using software simulation to produce radio propagation map

As most WLAN are working in an indoor environment, the transmission path between the transmitter and receiver can be severely obstructed by walls, furniture and even people, reflecting and diffracting signals make the radio propagation indoors even more unpredictable. Although a well planned site survey can produce a radio propagation map for an indoor environment, it is a time consuming and laborious approach. Another alternative method is to simulate the indoor propagation using software and map data (e.g. floor plans).

Wireless InSite™ [125] is a powerful tool for modelling the effects of buildings and terrain on the propagation of electromagnetic waves. It can be used to predict how the locations of transmitters and receivers within an urban area will affect the signal strength. In an indoor environment, InSite can predict received signal strength anywhere using its 3D ray tracing models. The ray tracing model sums up every signal from the transmitter to receiver including line of sight transmitted, reflected and diffracted in a full 3D environment. This method gives a big advantage over traditional probability based models.

By setting up the floor plan (walls, floors, ceilings, obstacles etc), and the radio transmitter (e.g. access point) with a proper antenna, the software can calculate the radio propagation map in a 3-D environment. The output file is used to initialise the WLAN simulator.

5.6 Applying radiation pattern from a semi-smart antenna

InSite supports various standard antennas including the popular dipole, monopole,
aperture antennas as well as user defined antennas. Here, the propagation pattern of the semi-smart antenna can be simulated with InSite to produce an initial radio propagation map.

It is possible to define a directional antenna by the main lobe or main beam of a high gain antenna while disregarding the sidelobes [126].

The angle of the slanting edge of the antenna with the horizontal is approximately 45°. The angle of radiation towards the likely region of users is approximately 60°. This corresponds to a requirement of a 3dB beamwidth of 60°.

By defining the required parameters (beamwidth 90°, 3dB width 60°) in Wireless Insite, the radio propagation pattern of one face of the antenna can be determined.

Figure 5.6: Simple definition of a directional antenna.

Figure 5.7: Setting up the one-face directional antenna in InSite.
One transmitter was set up with one directional antenna as mentioned above. The transmitter was placed 10m high with a 45 degree decline towards the floor as initially designed. A grid of receivers with omni-directional antennas were evenly distributed over the whole study area (a large hall with several doors and windows on the walls).

The power of each face of the semi-smart antenna is controlled by the digital step attenuator with 20 steps (0-20 dB). The power from the AP is set to be 20 dBm (100 mW, which is the maximum radiation power of 802.11 allowed in most countries of the world). So that the range of power control for each face of the antenna is 0 to 20 dBm (1 mW to 100 mW).

5.7 Simulating a Pseudo Real Scenario

Building a “real” environment involves building a scenario that includes the floor plan, material used, location of transmitter and receivers, antenna types for transmitters and receivers. Once this has been specified, InSite can calculate the signal strength throughout the location.

The test scenario used for this research is a large hall with four 802.11b access points at each corner of the hall, which is 90x90m with several doors and windows on the walls and different obstacles within the hall. This hall can be seen as a large and empty environment such as an airport lounge.

The four AP transmitters were set up with one-face directional antennas (Figure 5.8). Figure 5.8 illustrates the environment layout and the positions and directions of the APs.
InSite simulation generates a signal strength power map (Figure 5.9) illustrated by colour (red to blue represent signal strength from high to low). From the figure it can be seen that the radio propagation is not evenly distributed because of multi-path effect caused by walls etc.

Figure 5.8: Simulation environment and position of APs

Figure 5.9: Signal strength map of a one-face semi-smart antenna put at a corner of the hall.
The simulator takes the result files produced by Wireless Insite to produce its own signal strength map in polar coordinates with each QC being given an average signal strength reading based on the average of the sample points inside the QC area. With this map, each AP agent can get its own view of the multi-path rich environment as the basis of the later negotiations. Figure 5.10 shows the QC map of an AP, different colours representing the average signal power of the QC area; this represent an approximation of the radio coverage of the AP in Figure 5.9.

![Figure 5.10: QC representation of one AP's coverage, different colour represents different signal power.](image)

Simulations were performed on this geographical scenario with increasing user traffic from up to 96 IEEE 802.11b, 20% of which are uniformly distributed with the other 80% forming a hotspot near the upper left corner of the hall. The users in the hot spot follow a uniform distribution for both x and y coordinates. Figure 5.11 illustrates the user distribution in the simulation environment (each dot represents a single user).
Figure 5.11: distribution scenario

The following chart summarises the simulation setup:

<table>
<thead>
<tr>
<th>Item:</th>
<th>Setup:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation area:</td>
<td>90x90 m hall</td>
</tr>
<tr>
<td>Number of APs:</td>
<td>4</td>
</tr>
<tr>
<td>Antenna of AP:</td>
<td>Each AP is equipped with a one face of the semi-smart antenna, power control is range from 0 to 20 dBm (1 – 100 mW) with 20 step controls. Initial power is set to 10 dBm (10 mW)</td>
</tr>
<tr>
<td>Total number of users</td>
<td>Up to 96</td>
</tr>
<tr>
<td>User distribution:</td>
<td>Continually increasing users, 20% of them uniformly distributed. 80% forming a hot spot (the locations of the users in hot spot follow normal distributions)</td>
</tr>
<tr>
<td>Hot spot central position:</td>
<td>X: 30 m, Y: 30 m from the up-left corner (mean value of normal distributions)</td>
</tr>
<tr>
<td>Hot spot radius:</td>
<td>10 m (standard deviation of the normal distributions)</td>
</tr>
<tr>
<td>Wireless standard:</td>
<td>IEEE 802.11b</td>
</tr>
<tr>
<td>Maximum net throughput (one AP)</td>
<td>6.8 Mbit/s</td>
</tr>
<tr>
<td>User traffic demand:</td>
<td>Constant bit rate:, each STA's traffic demand range: 0 -2.5 Mbit/s, beta distributed with mean: 1.25 Mbit/s</td>
</tr>
</tbody>
</table>

Table 5.2: Simulation setup for modified negotiation algorithm.

As the number of users increases in the simulation environment, the heavily loaded AP
will start the negotiation process when its utilisation is above a certain threshold. Lightly loaded neighbouring APs may help by mutually changing the radiation patterns. Those STAs involved will then be forced to de-associate from the heavily loaded AP and search for a new AP with better signal strength. The non-uniform geographically distributed traffic triggers negotiation among the 4 AP as expected.

The system performance is evaluated by the same metric used in Chapter 4: the data rate to users (expressed here as overall system load/demand). When load/demand is 1, it means the system can serve the demand from all users. When this value is less than 1, one or more AP is experiencing overload and cannot serve the traffic demand).

In order to minimise the random effects of individual simulations, multiple simulation runs using different random seeds were performed and 20 independent experiments were averaged to produce the results shown in Figure 5.12. Confidence bars are added to the curves showing each individual simulation’s maximum and minimum value.

![Graph showing system performance (Load/Demand) average of 20 simulation runs](image)

**Figure 5.12: System performance (Load/Demand) average of 20 simulation runs**

The results show the same tendencies as those in Chapter 4 with the load balancing maintaining full throughput for a greater number of users, but as more users are added, the total demand is more than can be supported so the load/demand ratio is less than 1.
The two curves eventually start to merge, which means that all APs are fully loaded for both traditional network and negotiation. In that situation, there is no way to increase system capacity except by installing new APs.

The improvement curve of system’s load/demand can be shown in Figure 5.13. Using agent-based negotiation and semi-smart antenna radio pattern control in this case can improve the overall system performance up to 32%. This figure shows other features:

- The system can handle approximately double the number of users at full capacity, i.e. with the carried load being the same as demand. (from 10 users for traditional network to 20 users for network with negotiation).
- With greater congestion the agent-based system cannot carry all the demand, but the improvement continues to increase: when there are about 30 users, the improvement reached the peak value of about 32%.
- As the congestion grows the benefit starts to tail off, as was seen in Chapter 4, since there is finite capacity available from APs. In this case, however, the decline is faster as there are only 4 APs compared with 16 in the previous chapter.

In a real environment, like this simulation setup, the number of users would not normally reach as much as 100 users and 20-40 would be seen as reasonably high usage of the network. In such a situation, negotiation can greatly improve the system capacity.
Since the confidence bars in the previous results overlap, it is worth showing the result from a single simulation result to show that individual results benefit – this is shown in Figure 5.14 which demonstrates that an individual simulation does show better performance with negotiation.

**Figure 5.13: Improvement obtained for Load/Demand**

**Figure 5.14: System performance (Load/Demand) of single simulation run.**
From the results we can see that with agent-based cooperative negotiation and semi-smart antennas, a WLAN network can support more high-demand users without adding extra APs or changing the location of those APs.

The system performance can be evaluated with other metrics as well as load/demand. One of particular interest is network resource utilisation: in this measure, the total network utilisation is measured in order to evaluate the usage of network capacity. The network utilisation is defined here as the total load of the system (4 APs) divided by the total available capacity (4 AP’s potential capacity). The results are shown in Figure 5.15, which shows the improvement of using negotiations over a traditional network.

![Network Utilisation Improvement](image)

**Figure 5.15: network utilisation improvement**

### 5.8 Obstacles in an indoor environment

The previous environment was a large hall with a wall and room inside it. In order to test the effect of obstacles introduced to the network, it is worthwhile to run simulations without interior obstacles and compare the results with those of the previous scenario.
The interior walls are removed and the same simulations are performed (20 simulation runs again) and the comparison is shown in Figure 5.16 which shows that the performance degradation with obstacles is worse with the conventional network than with the negotiation.

This is because the obstacles in the big hall produce propagation shadows, and STAs in that shadow area will get a degraded signal and data rate. However, using negotiation and semi-smart antenna gives a performance that is very close to that without obstacles; as a result the improvement of system load/demand using negotiation is much larger for the environment with obstacles (Figure 5.17). This interesting result shows that with negotiation and semi-smart antennas, the improvement of performance is even better in a multi-path rich environment.

Figure 5.16: System performance (load/demand) for different environments
Figure 5.17: Improvement of load/demand for different environments.

5.9 Summary

This chapter enhanced the previous agent-based negotiation approach in order to test it in an indoor constrained environment.

Adaptive data rate support was added to take into account the fact that an 802.11 station can dynamically adjust its radio modulation according to the single strength and quality, so that the maximum data rate achievable is changed as well. An empirical data rate model (maximum data rate vs. SNR) was used in order to produce more realistic results.

The core negotiation algorithm was also modified in order to adapt to real network deployment where APs’ coverage is largely overlapping and the cell “border” concept no longer exists. The key modification is that AP agents need to use a de-association technique and mutually changing radiation patterns in order to shift mobile users from heavily loaded APs to lightly loaded APs.

In order to test the performance of the negotiation algorithm in a multi-path
constrained environment, a realistic test environment was simulated. Signal propagations calculations were performed using Wireless InSite in order to achieve most accurate results in a multi-path environment and simulation results show that even when only 4 APs available, using negotiation and semi-smart antennas, the system capacity (in terms of overall load/demand) can still improve significantly and the effect is most noticeable in bad multi-path environments with obstacles.
Chapter 6  Global optimisation algorithm

6.1 Introduction

The work in the previous chapters was inspired by distributed control of cellular networks, but WLANs are very different – especially in an indoor environment. WLANs do not have the scale and number of users that a GSM or 3G network does, so that more central collection of information can be used. In this chapter, a novel global optimisation algorithm (GOA) is implemented that uses centralised controls instead of negotiation between agents. The detailed algorithm, simulation and results comparison are analysed.

The GOA is developed as an optimisation technique to address the load balancing problem for WLAN in an indoor environment. Instead of quantisation cells representing the AP’s coverage and the minimum processing unit for negotiation, mobile stations are treated individually. With the GOA, APs periodically calculate their load utility and when the utility is higher than a threshold, the heavily loaded AP calculates and finds the least important STA(s) and informs the global optimiser that they will be unserved STA(s). The unserved STA(s) will attract adjacent APs to serve themselves and as a result, the STA(s) may be served by new AP so accomplishing the aim of load balancing. The process is iterated until all unserved STAs are re-associated or the iteration limit reached. During the iteration process, many power changes, de-associations and re-associations may occur, so the APs are only commanded to perform the antenna power change and actual de-associating of STAs at the final step.

The iteration process can be running at a central server; it communicates with all APs by SNMP (for collecting network utilisation etc.) and agent language (to control the antenna agent to adjust sector power). The GOA can be run as a software service on one of the servers on the network without adding any new hardware.

6.2 Global optimisation for constrained 802.11 WLANs

In the system level WLAN simulation, the load balancing problem can be formalised as follows.
Suppose that a WLAN network contains \( m \) APs and \( n \) STAs (mobile clients that are currently served by any of the APs in the WLAN). Each AP has some constraints, such as its maximum aggregated capacity, the maximum and minimum power of each antenna sector and the maximum distance it can serve a STA. The distance from a STA to its serving AP also needs to be minimised, in order to save battery power and reduce interference to others. The WLAN load balancing process is to maximise the network resource utilisation by allocating the \( n \) STAs equally over \( m \) APs according to the congregated traffic demand, while still satisfying all these constraints.

This can be formalised as an optimisation problem with constraints. The objective is to increase the system resource. The constraints are listed below:

1. The total served demand at any AP should be less than its capacity.
2. Each sector’s transmission power should be in the range of its maximum and minimum transmission power, at a pre-defined power step.
3. Each STA’s maximum data rate that can be achieved is no higher than the data rate computed by equations (5.4)(5.5)(5.6).
4. Only one AP can serve a STA at any given time.
5. The distance from an AP to any served STA should be shorter than the maximum distance that the AP can transmit signals.

More explicitly, this can be formulated as:

\[
\text{Maximise } \sum_{i=1}^{m} \sum_{j=1}^{n} c_j \cdot x_{ij} 
\]  

(6.1)

Where \( n \) is the number of mobile clients (STAs), and \( m \) is the number of APs. \( c_j \) represents the current traffic load of the \( j \)-th STA. \( x_{ij} \) is a binary variable that indicates if the \( j \)-th STA has been served by the \( i \)-th AP (\( x_{ij} = 1 \)) or not (\( x_{ij} = 0 \)).
These objectives are constrained by:

\[
\begin{align*}
\sum_{j=1}^{n} c_j \cdot x_{ij} & \leq b_i, \quad \forall i = 1, \ldots, m, \\
\text{Power}_{\text{min}} & \leq \text{Power}_k \leq \text{Power}_{\text{max}}, \quad \forall i = 1, \ldots, m, \quad \forall k = 1, \ldots, 4, \\
\sum_{i=1}^{m} x_{ij} & \leq M, \quad \forall j = 1, \ldots, n, \\
d_j \cdot x_{ij} & \leq d_{i,\text{max}}, \quad \forall x_{ij} = 1, \\
x_{ij} & \in \{0,1\}, \quad 1 \leq j \leq n, \quad 1 \leq i \leq m, \\
c_j & \leq t_j, \quad t_j = \min \{D_j, T_{n-\text{user},j}\}, \quad \forall j = 1, \ldots, n, \\
\end{align*}
\]  

(6.2)

\( b_i \) represents the \( i \)-th AP’s maximum capacity. \( \text{Power}_{\text{min}} \) and \( \text{Power}_{\text{max}} \) are the minimum and maximum power of the antenna sector. \( M \) is the maximum number of APs that can serve a STA; in the current 802.11 standard only one AP can serve a STA at any giving time so \( M = 1 \). \( d_j \) is the distance from the \( j \)-th STA to the \( i \)-th AP, and \( d_{i,\text{max}} \) is the maximum distance allowed by the \( i \)-th AP. The current traffic load of the \( j \)-th STA \( c_j \) is no greater than \( t_j \), which is smallest value of either the STA’s traffic demand \( D_j \), or the \( T_{n-\text{user},j} \) (achievable speed of \( j \)-th STA, when there are \( n \)-user in the serving AP). \( T_{n-\text{user},j} \) can be further defined by (6.3), which is defined in (5.4)(5.5)(5.6) and has been explained in section 5.2 of this thesis.

\[
\begin{align*}
T_{n-\text{user},j} &= T_{\text{user}}(\text{SNR}) - \left\{ \sum_{k \in S} \left[ T_{\text{loss-compute}}(\text{SNR}_k) + T_{\text{loss-hidden}}(C_i, C_k) \right] \right\} - T_{\text{offset}} \\
T_{\text{loss-compute}}(\text{SNR}_k) &= A_{\text{compute}} \times [T_{\text{max}} - T_{\text{user}}(\text{SNR}_k)] \\
T_{\text{loss-hidden}}(C_i, C_k) &= A_{\text{hidden}} \times PL(C_i, C_k) \\
\end{align*}
\]  

(6.3)

This research proposes a new method to solve the WLAN load balancing with semi-smart antennas.

As a WLAN is not generally laid out in a hexagonal layout, it is inadequate to simply model the coverage area by hexagonal cells. In Figure 6.1, the cell frontier is defined as the maximum transmission outreach of its AP when the free space model applies (no
obstacles between AP and STA). Any mobile client outside the frontier is of no direct interest to the AP, as it cannot service any of it. Therefore, the fourth constraint described in formula (6.2) is satisfied automatically. Local polar coordinates \((r, \theta)\) are used for easy identifying serving STAs for the host AP, and global Cartesian coordinates \((x, y)\) are used for the absolute locations of STAs and communications between adjacent APs.

![Figure 6.1: Cell model for Global Optimisation Algorithm](image)

In the GOA, each serving STA was assigned a vector called repulsion force \(\mathbf{R}^0_{ij}\). The direction of the force is from the \(i\)-th traffic unit within its frontier and is expressed as in (6.4)

\[
\mathbf{R}^0_{ij} = (d_{ij, \text{max}} - d_{ij}, \theta_{ij})
\]  

(6.4)

Where \(d_{ij}\) and \(\theta_{ij}\) is the distance and angle from the \(i\)-th access point to the \(j\)-th mobile client, and \(d_{i, \text{max}}\) is the distance to the frontier for the \(i\)-th access point.

The radial repulsion force represents how interested the access point is in having the specific mobile client assigned to it. As the repulsion force decreases with distance, the mobile client closest to the access point has the higher priority to be served.
An AP’s serving range is set in such a way that the magnitude of the repulsion forces of clients are larger or equal to a threshold value, $|\bar{R}_{ij}^0|$, the amount of traffic demand inside is less than or equal to its capacity, and all antenna sectors radiation power is in the adjustable range. Thus the first and second constraints in (6.2) are enforced during the process of calculating access points serving contours.

Calculating the threshold $|\bar{R}_{ij}^0|$ is a local optimisation problem. The local list of mobile clients (i.e. clients within the frontier) is sorted by their repulsion forces into a descending order. The threshold is determined by the repulsion force magnitude of the last one that still satisfies all constraints. This local optimisation process is described in the flow chart in Figure 6.2. As the length of this local list is reasonably short, this simple searching algorithm is efficient enough for the need of computation performance. With light traffic load, all local traffic demand is well under the threshold so there is no need to readjust the list among the access points. However, the local lists of STAs will start adjustment as soon as one access point exceeds its maximum load.

Once the list of an AP has exceeded the threshold, which means the associated mobile clients’ demand are higher than the AP’s capacity, the AP will send details of unserved clients to the central optimiser. The central optimiser will then start a series of re-allocating iterations in order to reduce the number of unserved mobile clients. The reallocation of clients is achieved by adding attraction forces from the unserved mobile clients to all nearby served clients in order to adjust the initial repulsion forces. The local optimisations are re-performed with the updated repulsion forces. This process is iterated until the number of unserved clients reduces to zero or the pre-defined number of iterations (10 in the simulation) is reached. This is shown as the main loop in Figure 6.2.
One AP exceed max load, start

- Reached the maximum number of iteration?
  - Y
  - N

  - Optimization converged?
    - Y
    - N

    Get next AP to process

    Calculate repulsion forces for all STAs within coverage, merge with attraction forces if any.

    Sort the repulsion forces in descending order, and create a sorted list of STAs accordingly

    Find the last STA that its magnitude satisfy all constrains, and calculate the sector power

    Pass un-served STAs to best candidate helper AP

    Any un-served STAs

    - Y
      - Find the last STA that its magnitude satisfy all constrains, and calculate the sector power

    - N
      - More APs to process

End process

Figure 6.2: Flow chart for the process of the Global Optimisation Algorithm
In this manner, the APs try to serve the unserved STA as it distorts the attraction forces nearby, and finally the optimum solution can be obtained for the current traffic condition. Figure 6.3 shows an example of a scenario with three APs.

Figure 6.3: Three access points and an unserved mobile client

Suppose there is a traffic hot spot in the area of AP1 coverage; AP1 experiences high traffic load and mobile client V4 is exceeding the threshold. AP1 will then inform the central optimiser that V4 is an unserved mobile client. As the iteration is performed by APs in a one-by-one manner, AP2 will then check if V4 is in its coverage range. As AP2 can serve V4, it will then calculate attraction forces to all clients it currently serves (i.e. T1, T2 and T3). The attraction force from the \( k \)-th unserved client \( V_k \) to the \( j \)-th mobile client \( T_j \) at the \( i \)-th base access point \( AP_i \) during the \( \lambda \)-th iteration is defined as 

\[
\vec{A}_{ij} = \begin{cases} 
(w_0, \theta_y), & k = j \\
(w_0, \theta_y - \theta_k), & k \neq j 
\end{cases}
\]  (6.5)
Where $\theta_{ik}$ and $\theta_{ij}$ are the angles from the $i$-th access point $AP_i$ to $V_k$ and $T_j$, respectively, and $w_0$ is a constant for controlling the attraction force to repulsion force ratio. For instance, the $A_{43}$ in Figure 6.3 represents the attraction force at the mobile client $T_3$ generated from the unserved units $V_4$.

If there is more than one unserved mobile client within the frontier, the resultant attraction force $\vec{A}_j^\lambda$ for the client $T_j$ is calculated by the vector sum of attraction forces from all the unserved clients $V_k$, expressed as (6.6).

$$\vec{A}_j^\lambda = \sum_{k=0}^{U_i} A_{ij}^\lambda$$  \hspace{1cm} (6.6)

where $U_i$ is the number of unserved mobile clients near the $i$-th access points.

The new repulsion force $\vec{R}_j^\lambda$ in the $\lambda$-th iteration is obtained by adding two parts together: the previous repulsion force, and the projection of the new resultant attraction force in the radial direction. It is formulated as (6.7) in polar coordinates.

$$\vec{R}_j^\lambda = (\vec{R}_j^{\lambda-1} + |\vec{A}_j^\lambda| \cdot \cos(\theta_j - \theta_{ij}), \theta_{ij})$$  \hspace{1cm} (6.7)

Where $\vec{R}_j^{\lambda-1}$ is the repulsion force from the previous control iteration and $\theta_{ij}$ is the angle of the resultant attraction force $\vec{A}_j^\lambda$.

As shown in Figure 6.3, the repulsion forces are increased by the attraction forces at the same side of the AP, and decreased by them at the opposite side. This results in the AP favouring serving the clients towards the unserved client $V_4$ (including $V_4$ itself). As $AP_2$ tends to serve clients towards $V_4$ (also the hot spot area), the client at the other side (e.g. $T_2$) will get less repulsion force from the AP and may be selected as an unserved STA for the next iteration (if $AP_2$ cannot serve all clients including $V_4$). In this fashion, a chained handover ($T_2$ being served by $AP_3$ in the next iteration) is achieved as expected. This is the main purpose of employing attraction forces here as it is used as a means of inter-AP communication. The advantage of this approach is that it can make the iterative adjustments more directional and accurate. This is the key
difference from the negotiation approach (where random clients are selected as removal STAs) where co-operative negotiations are used as inter-AP communication techniques.

This adjustment iteration is repeated until the optimisation has converged or the maximum number of iterations has been reached. After that, the traffic hot-spots are dispersed through the whole WLAN.

If there are still unserved mobile clients, the clients will be re-assigned to the AP with the strongest signal power. This is because in 802.11 networks, all mobile clients are served no matter what the traffic load conditions are in the APs. These unserved clients along with served clients will share the total bandwidth and every client will experience a data rate lower than the demand.

Only after the optimisation converges (all unserved traffic units have been eliminated) or the maximum number of iterations has been reached, are steps taken to adjust the actual powers in the antenna sectors and to re-associate mobile clients. This means the network does not change its current state until the optimisation is done, so the oscillation described above will not cause any disturbance to the current network.

6.3 Simulation results

6.3.1 Comparison with negotiation.

The same testing environment as in Chapter 5 was used first in order to compare the Global Optimisation Algorithm with the negotiation algorithm: the same geographical layout and the same use of Wireless InSite. Positions of access points and the number of users and traffic were also the same.

20 independent simulation runs were performed and the results averaged in order to reduce the effect of choosing random seeds. Figure 6.4 shows the results with confidence bars added. These results show that by using the GOA, the network has a slightly better performance compared with the negotiation algorithm.
The improvement of the data rates to users (here expressed as system load/demand) is shown in Figure 6.5. From the result, we can see that using the GOA also significantly improves the system’s capacity in terms of system load/demand. The improvement is even greater than using the negotiation algorithms with a peak improvement of 38% when there are about 30-40 users. The improvement from the GOA remains better than that with negotiation even as the improvement starts to decline as the APs become more heavily loaded.
Figure 6.5: Comparison of improvement of user data rates

The reason the GOA is superior to negotiation is that the underlying attraction and repulsing forces are efficiently attracting radiation patterns towards unserved mobile users. The vector attraction forces can more precisely steering antennas radiation power towards the hot spot area. As with the negotiation algorithm, the AP agent chooses its neighbouring APs to take some of the load. The selection of the best neighbour is a complex task when the APs are not laid out in hexagonal shape; with the GOA, the built in attracting mechanism can more precisely reallocate radio resources towards the hot spot areas than the negotiation approach.

Figure 6.6 shows the improvement in achieved load for 40 individual active users in the network. (This is sorted in ascending order of original traffic demand.) The average data rate improvement is 51.5% and ranges from 0% (where STAs are in light loaded APs) to up to 144% (in heavy loaded APs). This shows that by using GOA with semi-smart antennas and in a congested environment with heavy load, individual users can benefit from achieving much better data rates.
6.4 Different hotspot formation

The previous study supposed that the 80% of all STAs form hot spots, but it is worth considering other distributions. Four sets of simulations using different hot spot populations are performed (0%, 40% 80% 100% of users in the hotspot), the rest of the users being uniformly distributed in the study area. For each hotspot population category 20 independent simulation runs were performed and the results are shown in Figure 6.7. The improvement in the system data rate is shown in Figure 6.8. The figures show that the more users there are in a hot spot, the more benefit there is to be obtained from the GOA. This is to be expected as the algorithm is designed to cope with hot-spot scenarios. However, even with uniform user distributions, the GOA based system can still perform better (up to 15%) than a conventional network. The reasons for this are:

- For a traditional network, the mobile users and overall data demand cannot be exactly equally distributed over all APs even if all users are uniformly
distributed.

- For a traditional network, each AP’s coverage area is different so that APs that cover a larger area tend to get more heavily loaded when users are uniformly distributed.

From these results, a WLAN network without hot spots occurring can still benefit from improved system capacity.

![Graphs showing system performance under different hot spot populations](image)

**Figure 6.7: System performance under different hot spot populations.**
6.5 Complex multi-path constrained environment

In this section, the GOA algorithm is simulated in a real environment: the third floor of Macao Polytechnic Institute academic building. The building is a rectangular shape measuring 80×200m. It contains many classrooms surrounding 3 open squares. In the middle on the right there is a lecture theatre where large classes or seminars can be held. The floor plan can be seen as in Figure 6.9.
In order to simulate radio propagations under such a complex environment, a 3-D model of the floor plan was made (Figure 6.10). Doors and windows are also modelled in order to maximally approximate the real environment.

The locations of the APs are shown as in Figure 6.11. Eight APs are evenly placed in the testing environment, with 6 APs around the perimeter and 2 in the middle. All corner APs are equipped with 2 faces of the semi-smart antennas, each face covering 90 degrees. The 2 APs in the central area are equipped with 4 faces of the semi-smart
antennas. All semi-smart antennas have step power control that is the same as previous setups (0 – 20 dBm).

By using the intelligent radio resource management techniques, it is possible to minimise the need for costly site planning. The APs are placed evenly covering the whole floor without any obvious coverage holes. The dynamically changing radiation patterns can balance the traffic among APs automatically according to the current geographical traffic load.

![Figure 6.11: APs’ antennas directions and layout](image)

By using Wireless InSite, each AP’s initial radiation map can be calculated and used as the basic path loss calculation for simulations. Figure 6.12 shows two AP’s radiation power map. From these results, the multi-path effects can be clearly seen.
Figure 6.12: simulated radio propagation for two APs

With this test, the environment is much larger than the previous simulations so the maximum number of active WLAN users is set to 200. In addition, in order to test the algorithm in more complex user scenarios, 2 hot spots are generated in parallel in the testing area. Figure 6.13 shows the location of the two hot spots, the red circle showing the hot spot location and approximate range.
Figure 6.13: Users' scenario, forming 2 hot spots.

The simulation setup is summarised as below:

<table>
<thead>
<tr>
<th>Item:</th>
<th>Setup:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation area:</td>
<td>80x200 m 3rd floor of MPI academic building</td>
</tr>
<tr>
<td>Number of AP:</td>
<td>8</td>
</tr>
<tr>
<td>Antenna of AP:</td>
<td>2 APs are equipped with 4 faces of semi-smart antennas. 6 are equipped with 2 faces of semi-smart antennas. Power control is range from 0 to 20 dBm (1 – 100 mW) with 20 step controls. Initial power is set to 10 dBm (10 mW)</td>
</tr>
<tr>
<td>Total number of users</td>
<td>Up to 200 802.11b clients</td>
</tr>
<tr>
<td>User distribution:</td>
<td>Continually increasing users, 20% of them are uniformly distributed. 80% forming two hot spot: (the locations of the users in hot spot follow normal distributions)</td>
</tr>
<tr>
<td>Hot spots central position:</td>
<td>centred at x:y=60:40 and x:y=140:40</td>
</tr>
<tr>
<td>Hot spot radius:</td>
<td>20 m (standard deviation of the normal distributions)</td>
</tr>
<tr>
<td>Wireless standard:</td>
<td>IEEE 802.11b</td>
</tr>
<tr>
<td>Maximum throughput (one AP)</td>
<td>6.8 Mbit/s</td>
</tr>
<tr>
<td>User traffic demand:</td>
<td>CBR, each STA’s traffic demand range: 0 - 2.5 Mbit/s, beta distributed with mean: 1.25 Mbit/s</td>
</tr>
</tbody>
</table>

Table 6.1: Simulation setup

Again, 20 independent simulations using different random seeds are performed; the result of load/demand is shown in Figure 6.14. The results show that using GOA with semi-smart antennas, the system has much better performance than a traditional
network setup. When there are over 200 users in the testing environment, there is little benefit from using GOA as the network is then very heavily loaded. This figures also added a new curve called the theoretical maximum value; it can only be achieved when all STAs optimally associated to APs by maximising all APs’ utilisation. As users are not uniformly distributed (hot spots exist) and limitation of APs’ radiation power, it can not be achieved actually.

![Graph showing load/demand comparison](image)

**Figure 6.14: Simulation result for Load/Demand.**

The improvement of load/demand is depicted in Figure 6.15. From this it can be seen the gain reaches its peak value of 36% gain when there are about 80 users in the system. Again, the system can accommodate 40 users without degrading user’s traffic demand compared with 20 users using a traditional network setup.
As there are 8 APs in the simulation environment, in order to check the traffic conditions for individual APs under heavy user traffic, a new evaluation method is used. For each AP, the number of associated users and traffic demands are recorded and, in order to easily compare, the APs are sorted by traffic demand or user number in ascending order. Using this measurement, it is possible to justify how efficiently the GOA can balance user traffic among all APs.

Figure 6.16 shows individual AP’s aggregated user traffic demand when there are 80 active users in the system. From the diagram, it can be seen that by using GOA, the demand is efficiently re-balanced among different APs. Without using the GOA two APs in the hot-spot area are experiencing heavy loading (over twice as much as they could cope with) while others are lightly loaded. By using GOA, the two heavily loaded APs have their traffic demand efficiently shifted to the other 6 lightly loaded APs. The reason why traffic cannot be fully balanced has many causes, such as the location and size of the hot spots, the location of APs, and the limits in the adjustable antenna’s radiation power.
The number of users associated with each AP is shown in Figure 6.17, which confirms that users are shifted from heavily loaded APs to more lightly loaded ones.
In this chapter, a new global optimisation algorithm is proposed and evaluated. Instead of using distributed negotiation controls, a global optimisation algorithm is used. This method uses a centralised control where an iterative calculation calculates each AP’s antennas radiation power and association controls. The optimisation goal is reached when either all users’ traffic demand is satisfied or the maximum number of iteration is reached (choosing the iteration number will affect the trade-off between optimisation speed and quality). The simulation results show significant improvement over the negotiation approach.

It would be expected that a centralised approach would be better than the distributed negotiation because (i) it will have a more complete picture of the whole network and (ii) the repulsion/attraction concept allows targeting of radio resources more precisely. The disadvantage of a centralised approach is that there is often a communication overhead, but a WLAN deployment is generally sufficiently small for this to be
neglected.

The central server can be placed in the backbone network of the WLAN (usually a LAN network) and perform the calculation, periodically monitoring all APs traffic conditions and performing an optimisation calculation if certain threshold is reached. The choice of period gives a trade-off between network stability (minimising antennas’ power fluctuation and users’ re-association processes) and performance (quickest re-balancing).
Chapter 7  Conclusion and future work.

7.1 Conclusion

The thesis has proposed a novel radio resource management framework for IEEE 802.11 based WLANs. The key concept is to use a low cost semi-smart antenna and intelligent control algorithms to dynamically balance traffic load between APs. The proposed 4 sector semi-smart antenna can produce controllable coverage so that the co-operatively changing coverage patterns can effectively shift mobile clients from heavily loaded APs to lightly loaded ones. Two different intelligent algorithms have been proposed and tested:

- distributed intelligent agents with a cooperative negotiation techniques
- a more centralised approach with the Global Optimisation Algorithm

In both cases there is significant benefit and the test scenarios show that more users can be accommodated and that those users will get better data throughput.

While it could be argued that the antennas are more expensive and that extra capacity could be achieved with more APs, it is important to remember that IEEE 802.11 only provides three non-overlapping channels so avoiding interference from other APs is a problem with network deployment. In this approach the gains are provided without having to increase the number of APs so reducing the interference problem and making site deployment easier.

The GOA approach also benefits from there being no need to upgrade firmware (adding an intelligent agent) to the APs if a central server is used, so it can be applied to any SNMP enabled APs. The obvious disadvantage of GOA is that if the central server fails, the algorithm fails.

7.2 Further Work

Future work is planned to evaluate experimentally the proposed solutions by setting up a physical environment with prototype APs and semi-smart antennas.

However, more importantly, with the new 802.11 standards coming into use, the effect of optimising subject to the additional constraint of managing QoS also needs to be...
considered. While load balancing will help to maintain QoS by moving clients away from heavily loaded APs, even better improvement might be obtained for high priority services by taking the QoS explicitly into account. IEEE 802.11e offers QoS but this is achieved at the MAC layer rather than by network control. By ensuring that radiation patterns are shaped to cover the high-priority STAs (and in congestion, removing coverage from low-priority ones) there would be an additional mechanism to offer priority.

Another consideration is the use of fully-smart antennas. As antenna technology develops, fully-smart systems may become sufficiently cheap to envisage being applied in a WLAN. Additional work to compare the benefits between full-smart and semi-smart is desirable since then it would be possible to a cost-benefit analysis.
Author’s additional publications


References


1998.


