Automated Network Selection for Service Delivery across All-IP Heterogeneous Wireless Systems

by

FAROOQ BARI

M.Sc., Virginia Tech, Blacksburg, VA, USA, 1994
B.Sc., University of Engineering and Technology, Lahore, Pakistan, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES (ELECTRICAL AND COMPUTER ENGINEERING)

THE UNIVERSITY OF BRITISH COLUMBIA
August 2007
© Farooq Bari, 2007
Abstract

A crucial step in achieving service delivery in heterogeneous all-IP wireless systems is the selection of an appropriate delivery network. Simple network selection mechanisms as they exist today cannot be applied in this new service environment. Because of the revenue and service impacting nature of the problem, it has been considered critical by the communication industry to find a good solution. In our research we have investigated the scope of the problem and addressed both the architectural and algorithmic aspects of the problem.

We have proposed an architectural framework that is based on the network assisting the terminal in the decision about network selection. The proposed solution minimizes usage of the wireless links and would work with currently deployed infrastructure. It is scalable, flexible and supports roaming.

We developed a two step decision process for a network-assisted network selection mechanism that combines non-compensatory and compensatory multi-attribute decision making (MADM) algorithms. We have proposed enhancements to several MADM algorithms for their application to network selection. In order to deal with ranking abnormalities, we have proposed an improvement to the standard TOPSIS algorithm when it is applied to network selection by adopting an iterative approach. We have identified the need for support of non-monotonic utility for attributes used in network selection decision making and demonstrated that GRA is better suited for achieving this type of optimization objective. We have modified ELECTRE so that it can be used for network selection with attributes exhibiting non-monotonic utility. Our contributions in terms of modification to TOPSIS and ELECTRE and the usage of GRA with non-monotonic utility are not specific to the problem of network selection but these ideas can be used in other problems with similar requirements. We have developed a decision strategy for network selection in the absence of precise input information. The decision process proposed by us uses fuzzy techniques along with data prediction for the network selection decision process in certain cases. We have developed a new function, Confidence Level, and used it as a decision support tool along with the sensitivity of the service/subscription to the missing/imprecise information.
# Table of Content

ABSTARCT ................................................................. ii
TABLE OF CONTENT ...................................................... iii
LIST OF TABLES ............................................................ vii
LIST OF FIGURES ........................................................... viii
LIST OF ACRONYMS AND SYMBOLS ................................. xi
ACKNOWLEDGMENTS ...................................................... xiv
CO-AUTHORSHIP STATEMENT ........................................... xv

1 INTRODUCTION ......................................................... 1

1.1 PRIOR WORK .......................................................... 2
1.2 MOTIVES AND OBJECTIVES ........................................ 6
1.3 MAIN CONTRIBUTIONS .............................................. 7
1.4 ORGANIZATION OF THESIS ........................................ 11
1.5 REFERENCES .......................................................... 13

2 ARCHITECTURAL FRAMEWORK FOR NETWORK SELECTION IN HETEROGENEOUS WIRELESS SYSTEMS ................................................................. 16

2.1 INTRODUCTION ........................................................ 16
2.2 FACTORS IMPACTING NETWORK SELECTION .................. 18
  2.2.1 Access Network Characteristics ................................ 18
  2.2.2 Service Selection .................................................. 18
  2.2.3 Roaming Support .................................................. 18
2.3 BACKGROUND AND RELATED WORK ............................ 18
  2.3.1 Related Recent Standards Activities ........................ 19
2.4 PROPOSED ARCHITECTURE ........................................ 21
  2.4.1 Terminal Based Approach ....................................... 21
  2.4.2 Network Assisted Approach .................................... 22
  2.4.3 Using DCN to Collect and Provide Access Network Characteristics ................................................................. 26
  2.4.4 Using SAN for Service Selection ................................ 27
  2.4.5 Using Enhanced AAA for Roaming Support ................. 28
  2.4.6 Information Collection and Exchange ........................ 29
2.5 DEPLOYMENT SCENARIOS
2.5.1 DCN (IPF) and SAN residing in the AN .............................................. 31
2.5.2 DCN (IPF) and SAN residing in local information provider network .............. 32
2.5.3 DCN (IPF) and SAN residing in the HN ................................................. 33
2.6 POLICY-BASED NETWORK SELECTION .................................................. 35
2.6.1 Adding Standards-based Policy Nodes to Network Selection Framework ........ 36
2.6.2 Example of Policy-based Automated Network Selection .......................... 38
2.7 CONCLUSION ...................................................................................... 41
2.8 REFERENCES .................................................................................... 42

3 USE OF MADM ALGORITHMS FOR AUTOMATED NETWORK SELECTION IN A HETEROGENEOUS WIRELESS NETWORK ENVIRONMENT ..................................... 44
3.1 INTRODUCTION ................................................................................ 44
3.2 BACKGROUND AND RELATED WORK ............................................... 45
3.3 DECISION PROCESS IN NETWORK SELECTION ................................. 46
3.4 MULTI ATTRIBUTE DECISION MAKING (MADM) ................................. 49
3.4.1 Non-Compensatory MADM Algorithms ......................................... 49
3.4.1.1 Decreasing Inter-Technology Handoffs ........................................ 50
3.4.2 Compensatory MADM Algorithms ................................................. 51
3.4.2.1 Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) ................................................................. 52
3.4.2.2 Assignment of Attribute Weights .................................................. 54
3.4.3 Results of Use Case Scenarios ......................................................... 56
3.4.4 Sensitivity of Network Selection to Dynamic Attribute Values .......... 59
3.4.5 Network Selection with Managed IP Networks .................................. 60
3.4.6 Network Selection for Accessing Multiple Services Simultaneously .... 61
3.5 CONCLUSIONS ................................................................................. 62
3.6 REFERENCES .................................................................................... 63

4 ENHANCEMENTS TO MADM ALGORITHMS FOR USE IN AUTOMATED NETWORK SELECTION WHERE SYSTEM PARAMETERS ARE KNOWN ............................................. 65
4.1 USE OF ITERATIVE TOPSIS FOR NETWORK SELECTION IN HETEROGENEOUS WIRELESS ACCESS .............................................................. 66
4.1.1 TOPSIS .......................................................... 67
4.1.2 Ranking Abnormalities .................................... 69
  4.1.2.1 Factors contributing to ranking abnormality for TOPSIS .... 70
4.1.3 Iterative MADM Approach using TOPSIS .......... 71
  4.1.3.1 Results for use case scenario .......................... 73

4.2 UTILITY ASPECTS OF ATTRIBUTES IN APPLICATION OF MADM ALGORITHMS TO NETWORK SELECTION 75
  4.2.1 Deficiency in Current Studies ............................ 75
  4.2.2 Use of Non-monotonic Utility for Attributes in Network Selection .............................................. 76
  4.2.3 Comparison of MADM Algorithms for Use with Non-monotonic Utilities of Attributes .................... 78
  4.2.4 Theory of Grey Relational Space ....................... 80
  4.2.5 Application of GRA Adapted to Network Selection with Non-monotonic Utility .................... 80
  4.2.6 Evaluation of Using Non-monotonic Utility in a Heterogeneous Wireless Network Environment .... 83
    4.2.6.1 Setting up GRA for Network Selection .............. 84

4.3 APPLICATION OF ELECTRE TO NETWORK SELECTION IN HETEROGENEOUS WIRELESS ACCESS .......... 88
  4.3.1 Evaluating MADM Algorithms for Use in Network Selection ......................................................... 88
    4.3.1.1 Accuracy of the results obtained from an algorithm ................................................................. 88
    4.3.1.2 Appropriateness of applying the algorithm to the problem ...................................................... 89
  4.3.2 Application of Modified ELECTRE to Network Selection ............................................................. 90
    4.3.2.1 Approach 1 ................................................. 94
    4.3.2.2 Approach 2 ................................................ 95
  4.3.3 Evaluation of Modified ELECTRE ...................... 96

4.4 CONCLUSION .................................................. 102
4.5 REFERENCES ................................................. 104

5 APPLICATION OF DATA PREDICTION AND FUZZY TECHNIQUES WITH MADM-BASED AUTOMATED NETWORK SELECTION ................................................................. 106
  5.1 INTRODUCTION ............................................... 106
  5.2 USING NON-COMPENSATING MADM ALGORITHMS WITH INCOMPLETE INFORMATION .......................... 107
  5.3 USING COMPENSATING MADM ALGORITHMS WITH DATA PREDICTION AND FUZZY INPUT .................... 108
    5.3.1 Scenario 1 - Imprecise or Missing Information ................................................................. 108
      5.3.1.1 Data Estimation ........................................ 111
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.1.2</td>
<td>Compensating MADM Algorithm</td>
<td>115</td>
</tr>
<tr>
<td>5.3.1.3</td>
<td>Defuzzification</td>
<td>119</td>
</tr>
<tr>
<td>5.3.1.4</td>
<td>Confidence Level</td>
<td>124</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Scenario 2 - Network Types with Non Crisp Attributes</td>
<td>131</td>
</tr>
<tr>
<td>5.4</td>
<td>CONCLUSION</td>
<td>135</td>
</tr>
<tr>
<td>5.5</td>
<td>REFERENCES</td>
<td>136</td>
</tr>
<tr>
<td>6</td>
<td>CONCLUSIONS</td>
<td>138</td>
</tr>
<tr>
<td>6.1</td>
<td>SUMMARY OF CONTRIBUTIONS</td>
<td>138</td>
</tr>
<tr>
<td>6.2</td>
<td>FUTURE WORK</td>
<td>141</td>
</tr>
<tr>
<td>6.3</td>
<td>REFERENCES</td>
<td>143</td>
</tr>
</tbody>
</table>
List of Tables

Table 2-1 Comparison of terminal-based and network-assisted mechanisms ............................................. 24
Table 3-1 Attributes for use in network selection ....................................................................................... 48
Table 3-2 Attribute values for Scenarios 1, 2, 3 & 4 .................................................................................. 56
Table 3-3 Results on sensitivity analysis .................................................................................................. 59
Table 3-4 Attribute values for Scenarios 1, 2, 3 & 4 .................................................................................. 60
Table 4-1 Attribute values for the candidate networks .............................................................................. 73
Table 4-2 Results of Iterative approach for TOPSIS ................................................................................ 74
Table 4-3 Reference attribute values ....................................................................................................... 85
Table 4-4 Assignment of attribute weights .................................................................................................. 85
Table 4-5 Attribute Values for Scenarios under Consideration ................................................................. 97
Table 4-6 Reference attribute values for voice over IP, streaming and web browsing services .............. 97
Table 4-7 Ranking for networks using standard ELECTRE method with Alternative 2 ......................... 98
Table 4-8 Ranking for networks using modified ELECTRE method with Alternative 2 ......................... 98
Table 4-9 voice over IP service .................................................................................................................. 99
Table 4-10 Streaming service .................................................................................................................... 100
Table 4-11 Web browsing service ............................................................................................................. 101
Table 5-1 Attributes and their weight assignments for streaming media and web browsing services while using GRA ................................................................................................................................. 115
Table 5-2 Attribute values for the five networks under consideration ...................................................... 117
Table 5-3 Lookup tables used by the decision maker for Confidence Level calculations ...................... 130
Table 5-4 Attribute values for network alternatives under consideration ............................................... 132
List of Figures

Figure 1.1 Organization of thesis .......................................................... 11
Figure 2.1 The problem of network selection ....................................... 17
Figure 2.2 Decision process in network selection using proposed network-assisted mechanism ......................................................... 25
Figure 2.3 A heterogeneous wireless system with access networks using different technologies .......................................................... 25
Figure 2.4 Logical nodes involved in network selection .......................... 26
Figure 2.5 Architecture for network related information collection and exchange .......................................................... 30
Figure 2.6 QoS attributes used for network selection, admission control, handoffs, packet scheduling can be the same but will have different update rate and the level of accuracy ......................................................... 30
Figure 2.7 DCN (IPF) and SAN residing in the AN ................................. 32
Figure 2.8 DCN (IPF) and SAN residing in local information provider’s network .......................................................... 33
Figure 2.9 DCN (IPF) and SAN residing in HN ......................................... 34
Figure 2.10 High level architecture for a heterogeneous wireless system with user’s HN providing the services .............................. 37
Figure 2.11 Logical nodes involved in network selection .......................................................... 38
Figure 2.12 High level communication with policy components residing in the HN. The PDF is enhanced to include the functionality of decision making for network selection. The decision making algorithms that can be used are described in later chapters .......................................................... 39
Figure 2.13 Signaling and bearer paths once AN#2 has been selected. The bearer path is route optimized via a local breakout .......................................................... 40
Figure 3.1 Decision making process in network selection .......................................................... 49
Figure 3.2 By using non-compensatory MADM algorithm, different mobility profile settings can provide selection of a different network .......................................................... 51
Figure 3.3 Example of indifference curve for TOPSIS while using attributes of cost per byte and throughput .......................................................... 54
Figure 3.4 Results for scenarios 1 & 3 ......................................................... 57
Figure 3.5 Results for scenarios 2 & 4 ......................................................... 58
Figure 4.1 Representation of an MADM problem with n alternatives and m attributes .......................................................... 67
Figure 4.2 Use of TOPSIS for a MADM problem with 4 alternatives and 2 attributes .......................................................... 70
Figure 4.3 Decision making process in Iterative TOPSIS approach .......................................................... 72
Figure 4.4 Assignment of weights to attribute values for a Bronze user subscription in the home operator network.

Figure 4.5 Graphical representation of a simple decision making scenario with one attribute and two networks.

Figure 4.6 Three ways of using reference attribute values and attribute weights with GRA algorithm. Approach 3 is recommended in network selection.

Figure 4.7 Results for network selection for three possible configurations of GRA. Configuration 3 is preferred as it provides maximum flexibility to fine tuning the algorithm to optimization objectives.

Figure 4.8 Weights associated with attributes for different services.

Figure 5.1 Use of MADM with imprecise information can help narrow down the candidate service delivery network options.

Figure 5.2 Distributed nature of data collection points across different access types and autonomous domains makes harmonized, accurate and real time parameter information challenging.

Figure 5.3 Steps involved in proposed network selection mechanism with imprecise information.

Figure 5.4 Fuzzy Numbers.

Figure 5.5 Example graphs of packet Delay (D), Jitter (J) and Latency (L) values using Network Utilization (U). It assumes correlation represented by $D = 0.225\times U^2 + 10$, $J = 0.05\times U^2$, $L = 0.0019\times U^3$ with 0.8U - 1.1U in the network utilization range of 5% and 70%.

Figure 5.6 Results for streaming media services using GRA algorithm with fuzzy input and with imprecise input information.

Figure 5.7 Use of CoG for defuzzification of a TFN representing the GRC in a fuzzy implementation of GRA for streaming media services.

Figure 5.8 Results of GRA algorithm for streaming media service after removing Ntwk#5.

Figure 5.9 Results for web browsing service using GRA algorithm with fuzzy input information and imprecise input information.

Figure 5.10 Use of CoG for defuzzification of a TFN representing the GRC in a fuzzy implementation of GRA for web browsing services.

Figure 5.11 Trend graph for relationships between Confidence Level and common service types when a QoS related parameter with dynamic characteristics has unreliable value and therefore is being predicted.

Figure 5.12 Trend graph for relationship between Confidence Level and reliability of estimated data for some of the common service types.
Figure 5.13 Trend graph for relationship between Confidence Level and time since last attribute value was received

Figure 5.14 Trend graph for relationship between Confidence Level and degree of correlation between known and unknown attributes

Figure 5.15 An alternative approach of using Confidence Level in the decision process

Figure 5.16 Graphical representation of how network congestion during service delivery time can be related to parameters such as the total bandwidth, bandwidth allowed per user, current network utilization and rate of change of network utilization

Figure 5.17 Results of fuzzy GRA algorithm

Figure 5.18 Use of CoG for defuzzification of TFN representing GRCs in the fuzzy implementation of GRA
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3GPP</td>
<td>3rd Generation Partnership Program</td>
</tr>
<tr>
<td>AAA</td>
<td>Authentication, Authorization and Accounting</td>
</tr>
<tr>
<td>AB</td>
<td>Allowed Bandwidth</td>
</tr>
<tr>
<td>AHP</td>
<td>Analytic Hierarchy Process</td>
</tr>
<tr>
<td>AM</td>
<td>Authentication Mechanism</td>
</tr>
<tr>
<td>AN</td>
<td>Access Network</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>AT</td>
<td>Access Technology</td>
</tr>
<tr>
<td>BAS</td>
<td>Basic Access Signaling</td>
</tr>
<tr>
<td>CA</td>
<td>Coverage Area</td>
</tr>
<tr>
<td>CB</td>
<td>Cost per Byte</td>
</tr>
<tr>
<td>CALEA</td>
<td>Communications Assistance for Law Enforcement Act of 1994</td>
</tr>
<tr>
<td>CoG</td>
<td>Center of Gravity</td>
</tr>
<tr>
<td>CS</td>
<td>Command Services</td>
</tr>
<tr>
<td>CSCF</td>
<td>Call Session Control Function</td>
</tr>
<tr>
<td>D</td>
<td>Packet Delay</td>
</tr>
<tr>
<td>DCN</td>
<td>Data Collection Node</td>
</tr>
<tr>
<td>DNS</td>
<td>Domain Name Server</td>
</tr>
<tr>
<td>E911</td>
<td>Enhanced 911 (emergency service)</td>
</tr>
<tr>
<td>EAP</td>
<td>Extensible Authentication Protocol</td>
</tr>
<tr>
<td>ELECTRE</td>
<td>Elimination and Choice Translating Priority</td>
</tr>
<tr>
<td>ES</td>
<td>Event Services</td>
</tr>
<tr>
<td>GAS</td>
<td>Generic Advertisement Services</td>
</tr>
<tr>
<td>GL</td>
<td>Geographic Location</td>
</tr>
<tr>
<td>GRA</td>
<td>Grey Relational Analysis</td>
</tr>
<tr>
<td>GRC</td>
<td>Grey Relational Coefficient</td>
</tr>
<tr>
<td>GRS</td>
<td>Grey Relational Space</td>
</tr>
<tr>
<td>GSM</td>
<td>Groupe Spécial Mobile</td>
</tr>
<tr>
<td>HN</td>
<td>Home Network</td>
</tr>
<tr>
<td>HPLMN</td>
<td>Home Public Land Mobile Network</td>
</tr>
<tr>
<td>ICF</td>
<td>Information Collection Function</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>IMS</td>
<td>Internet Multimedia Subsystem</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>IPF</td>
<td>Information Provider Function</td>
</tr>
<tr>
<td>IS</td>
<td>Information Services</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronic Engineers</td>
</tr>
<tr>
<td>IETF</td>
<td>Internet Engineering Task Force</td>
</tr>
<tr>
<td>IRTF</td>
<td>Internet Research Task Force</td>
</tr>
<tr>
<td>J</td>
<td>Packet Jitter</td>
</tr>
<tr>
<td>L</td>
<td>Packet Loss</td>
</tr>
<tr>
<td>LDAP</td>
<td>Lightweight Directory Access Protocol</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MADM</td>
<td>Multi Attribute Decision Making</td>
</tr>
<tr>
<td>MIH</td>
<td>Media Independent Handover</td>
</tr>
<tr>
<td>NAI</td>
<td>Network Access Identifier</td>
</tr>
<tr>
<td>ON</td>
<td>Operator Name</td>
</tr>
<tr>
<td>PDF</td>
<td>Policy Decision Function</td>
</tr>
<tr>
<td>PHY</td>
<td>Physical Layer</td>
</tr>
<tr>
<td>PEF</td>
<td>Policy Enforcement Function</td>
</tr>
<tr>
<td>PLMN</td>
<td>Public Land Mobile Network</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RADIUS</td>
<td>Remote Authentication Dial-In User Server/Service</td>
</tr>
<tr>
<td>RFC</td>
<td>Request For Comments</td>
</tr>
<tr>
<td>SA</td>
<td>Services Available</td>
</tr>
<tr>
<td>SAN</td>
<td>Service Announcement Node</td>
</tr>
<tr>
<td>SAE</td>
<td>System Architecture Evolution</td>
</tr>
<tr>
<td>SAW</td>
<td>Simple Additive Weighing</td>
</tr>
<tr>
<td>SIM</td>
<td>Subscriber Identification Module</td>
</tr>
<tr>
<td>SLA</td>
<td>Service Level Agreement</td>
</tr>
<tr>
<td>SLP</td>
<td>Service Location Protocol</td>
</tr>
<tr>
<td>SSID</td>
<td>Service Set Identifier</td>
</tr>
<tr>
<td>TB</td>
<td>Total Bandwidth</td>
</tr>
<tr>
<td>TOPSIS</td>
<td>Technique for Order Preference by Similarity to Ideal Solution</td>
</tr>
<tr>
<td>U</td>
<td>Network Utilization</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunication System</td>
</tr>
<tr>
<td>VoIP</td>
<td>Voice over IP</td>
</tr>
<tr>
<td>VPLMN</td>
<td>Visited Public Land Mobile Network</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
<tr>
<td>WPM</td>
<td>Weighed Product Model</td>
</tr>
<tr>
<td>WSM</td>
<td>Weighed Sum Model</td>
</tr>
<tr>
<td>WWAN</td>
<td>Wireless Wide Area Network</td>
</tr>
</tbody>
</table>
Acknowledgments

I would like to thank Dr. Victor Leung for his guidance during my stay at UBC. As my graduate advisor he provided me with invaluable suggestions. His insightful comments while reviewing my work not only helped me in my research but also allowed me to improve my professional skills as a researcher.

I would like to thank Professor Panos Nasiopoulos and Professor Son T. Vuong for serving on my dissertation committee.
Co-authorship Statement

I am the first author and principal contributor of all manuscript chapters. All manuscript chapters are co-authored with Dr. V.C.M. Leung who supervised the thesis research.
1 Introduction

In the past few decades several wide area and local area wireless access technologies have emerged. While many of these technologies have been successfully deployed and continued to evolve, none of them provides a universal coverage to cater to the mobile lifestyle of today’s users. The desire and expectation to have ubiquitous broadband connectivity all the time and the emergence of multimode devices are forcing the network operators to look at new and innovative ways of service delivery including the possibility of interworking of these evolving access technologies. An example of a multimode device is a terminal that supports Institute of Electrical and Electronics Engineers (IEEE) 802.11 Wireless Local Area Network (WLAN) technology and Groupe Special Mobile (GSM) Wireless Wide Area Network (WWAN) technology. A major challenge in this new environment is network selection, i.e., identifying the best suited service delivery network when the user has multiple networks to choose from. Network selection becomes especially challenging for multimode devices that have the option to use different all-IP wireless access technologies to gain access to various services. To meet the challenge of ubiquitous service delivery, the network selection mechanism should also work under roaming and when there are more than one authentication mechanisms and credentials available to access the available networks. For example, a user may have multiple accounts with the same or different service providers, such as the user’s own personal and job related accounts. In order to get better coverage the user may also access different networks using more than one types of authentication credential, e.g., Subscriber Identification Module (SIM) based or user ID/password. These credentials can map to the same or different accounts that can be accessed for delivering a variety of services over a number of access networks with different Quality of Service (QoS) capabilities or conditions.

In the case of current GSM systems for example, network selection involves a scan for the network identities, i.e., Public Land Mobile Network IDs (PLMN IDs), followed by a selection of one of them based on the pre-provisioned information in the terminal about preferred PLMN IDs and forbidden PLMN IDs. In the case of IEEE 802.11 WLANs, the beacon and probe request/response mechanism provides a way for stations to discover Access Points (APs) using Service Set Identifiers (SSIDs). Based on this information and signal strength, the station can decide which AP to associate with. Such simplistic approaches, however, are unlikely to yield an optimal network selection in interworked heterogeneous all-IP systems that use different
access technologies and authentication mechanisms while delivering a range of services from a variety of operators.

To deliver a good customer experience, selection of an optimal network is essential. The decision by the user / terminal to select an Access Network (AN) and attach to it can be either manual or automatic. In general, in order to have a good customer experience, automatic network selection mechanisms would be preferred with manual selection kept as a possible fallback. Manual network selection is generally quite simple in that it typically comprises the user manually picking a network from a list of networks visible to the terminal. Manual selection can fail for a variety of reasons such as the user's home operator not having roaming agreements with the chosen network operator. Automatic network selection could be based on much more information that would allow a more informed decision to be made. The type of information, its collection and its usage in the decision process all are important aspects of ongoing research. The information used can include the service being requested and the service supported, network identity and its roaming agreements, authentication mechanisms supported, QoS provided, cost per byte, etc. Many of these information elements are static while others are dynamic. It is also important to define an architecture with details about the entities involved, the information flows, and the processes employed for network selection. As described later in the thesis the possibility of network-based, terminal-based and network-assisted network selection exist. Depending upon the solution proposed the information may have to be collected, network selection decision made and conveyed before the terminal gets authenticated.

1.1 Prior Work

Architectural aspects of network selection for interworked all-IP heterogeneous wireless networks operating in autonomous domains have so far been a largely unexplored area of research. In [1], system level requirements such as the capability to discover all the radio networks using different technologies available in an area have been discussed. A proof of concept system that uses Basic Access Signaling (BAS) to identify the radio technologies has been proposed for this purpose. The paper however does not consider realistic scenarios based on current network deployments. For example, it ignores the fact that networks are typically operated and managed by autonomous business entities that can use different technology standards as well as different authentication mechanisms, roaming arrangements, and billing systems. The architecture also does not seem to be based on any current business model in which
a user has a subscription/billing arrangement with his Home Network (HN) operator for delivery of services that the user may want to access.

Independent of the architectural aspects, there have been some studies done in the past for algorithmic schemes to select the appropriate service delivery network. In [2], the use of analytic hierarchy process with Grey system theory for network selection between Universal Mobile Telecommunications Systems (UMTSs) and WLANs has been proposed. In [3], network selection on a per call basis based on current market conditions such as the cost has been described. In [4], a resource allocation scheme in a heterogeneous environment has been proposed. A basic fuzzy logic solution for access network selection in a heterogeneous networking environment has been described in [5]. A simple policy-enabled handoff system across heterogeneous wireless networks is presented in [6]. These papers however do not completely address the complex problem of network selection. Rather simplistic assumptions have been made due to a lack of understanding of the scope of the problem and unavailability of an architectural framework. The factors considered in the decision processes are in many cases insufficient; e.g., information about access types supported, authentication types supported, roaming partners supported has not been considered in the decision processes described in these papers. The analysis on the choice of the decision making algorithm has been limited when it comes to their suitability for application to the problem of network selection. They also do not provide or refer to a viable architectural framework in which the selection mechanism can work and hence they fail to provide a complete and deployable solution to the problem.

Aspects of the problem have been discussed in various industry standardization forums. Here we provide a snapshot of the key related activities that have occurred in the Internet Engineering Task Force (IETF), IEEE and 3rd Generation Partnership Program (3GPP) forums. In IETF, several Working Groups have looked into the network selection problem. The IETF work has dealt mainly with the problem from an Authentication, Authorization and Accounting (AAA) perspective in a homogeneous system. Much of this work was done before the entire scope of the problem was evident for today’s all-IP heterogeneous wireless networking environment. Nonetheless the work done by these groups is useful in proposing a comprehensive solution to the problem. For example, the ROAMOPS WG defined the Network Access Identifier (NAI) originally in RFC 2486 [7], and subsequently updated it in RFC 4282 [8]. NAI is useful in providing user and HN identity in a standard and interoperable manner to the network entities. RFC 2194 [9] describes the AAA aspects of some roaming implementations while RFC 2607
describes the use of proxies in roaming. The CARD protocol defined in RFC 4066 [11] can assist in discovery of suitable APs. RFC 3280 [12] developed by the Public Key Infrastructure X.509 (PKIX) WG addresses issues of certificate selection for use in authentication. The AAA WG also developed some service discovery mechanisms within Diameter's RFC 3588 [13]. In RFC 4284 [14] Extensible Authentication Protocol (EAP)-Request/Identity has been used to provide "realm hints" useful for the selection of user identities. The technique was used by 3GPP in advertisement of WLANs interworking with cellular systems [15]. However, advertisement mechanisms based on the use of EAP-Request/Identity have been found to have scalability problems with a substantial negative impact on the handoff latency if such mechanisms are used for the selection of candidate networks during handoffs [16].

IEEE 802 working groups have also worked on the problem of network selection for WLANs that use the IEEE 802.11 technology. [17] describes the beacon and probe response mechanisms in IEEE 802.11 based systems. While beacons can potentially be used to advertise more than one networks, they may be sent only at a rate within the base rate set, which typically consists of the lowest supported rate. Studies [18] have shown Medium Access Control (MAC) layer performance problems with this approach and [19] has identified scaling issues if the beacon interval is lowered. [20] discusses how authentication models in WLANs can be migrated from existing models to new ones that use EAP layer indications or by intelligent use of SSIDs properly structured to represent more than the local network. It also identified scalability issues with virtual APs. [21] presents mechanisms currently used to provide virtual AP capabilities within a single physical AP that would allow multiple networks to be advertised from one AP. A virtual AP appears at the MAC and IP layers to be a distinct physical AP. Compatibility with existing 802.11 station implementations is possible if each virtual AP uses a distinct MAC address (BSSID) for use in beacons and probe responses. However simulation results presented in [19] have highlighted a scalability problem with this approach with a limit of approximately 20 virtual APs per physical AP. Work on network selection is ongoing in IEEE 80211u and IEEE 802.21. The author of this thesis has contributed to this work [23][24] during the research. More detail on it is provided in Chapter 2 where an architectural framework for solving the network selection problem is described.

3GPP cellular network selection, also known as PLMN selection, is described in [25] where the terminal monitors surrounding cells and prioritizes them based on signal strengths before selecting a new potential target cell. Each cell broadcasts its PLMN ID. A terminal, provisioned
with a mapping of operator to its PLMN ID, may automatically select cells that belong to its Home PLMN (HPLMN), Registered PLMN or an allowed set of Visited PLMNs (VPLMNs). For GSM systems, the PLMN ID lists are prioritized and stored in the SIM. For manual PLMN selection, the terminal collects the PLMN IDs it learns from the surrounding cells and enables the user to choose the desired PLMN based on mapping of these IDs to operator names. After the PLMN has been selected, cell prioritization takes place, in order to select the appropriate target cell. TS 23.234 [15] describes interworking of WLANs with 2nd Generation (2G) and 3rd Generation (3G) cellular networks and discusses realm discovery and network selection issues in that environment. In the WLAN roaming scenarios considered by 3GPP, multiple network levels may be present, and the hotspot owner may have a contract with a provider who in turn has a contract with a 3G network, which may have a roaming agreement with other networks. The specification requires that network discovery be performed as specified in the relevant WLAN link layer standards. In addition to network discovery, as part of network selection it is necessary to select an intermediary roaming partner or broker. In 3GPP, the intermediary networks are also assumed to be PLMNs, and it is assumed that an AN may have a roaming agreement with more than one PLMNs. The PLMN may be a HPLMN or a VPLMN, where roaming is supported. GSM/UMTS roaming principles are employed for routing AAA requests from the VPLMN to the HPLMN using either RADIUS or Diameter. The roaming partner network related information provided via EAP methods by the visited network as proposed in RFC 4284 and used in 3GPP is only considered a hint and typically provided when a network considered to be unreachable via roaming agreements is encountered. The station can also manually trigger a request for this information by including an unknown realm (known as the Alternative NAI) within the EAP-Response/Identity. A realm guaranteed not to be reachable within 3GPP networks is utilized for this purpose. The security requirements for PLMN selection are discussed in a 3GPP contribution [26], which concludes that both SSID and EAP-based mechanisms have similar security weaknesses. As a result, it recommends that PLMN advertisements be considered only as hints.

It is clear from the discussion above that while various standardization forums are trying to address the problem of network selection, the scope of the work and the solutions provided are limited or restricted by the mandate of the forum. For example, 3GPP has only considered the use of SIM based authentication and involvement of PLMN operators for roaming scenarios. Other independent studies have also discussed the need for network selection where an AN may
have more than one roaming relationships to a HN; none of them however provides a comprehensive approach towards solving the problem. [27] describes solutions to the realm selection problem based on EAP, SSID and PEAP-based mechanisms. [28] discusses the realm and network selection problem, concentrating primarily on the discovery of ANs meeting a set of criteria.

1.2 Motives and Objectives

The objective of this research has been to provide solutions to key issues associated with the process of network selection for service delivery over a heterogeneous mix of all-IP wireless networks, possibly managed by more than one network operators. In this context, a key area of work was the development of a common architectural framework for network selection across heterogeneous wireless networks.

It is well known that as the number of factors influencing a decision increases from more than a few, it becomes difficult for humans to directly compare the alternatives and hence the need for decision tools. In the thesis, we have focused our research of decision making algorithms on the application of various Multi Attribute Decision Making (MADM) algorithms, as they have been successfully applied to real life decision problems in fields such as operation research for several decades. Such algorithms are designed to take into consideration multiple input factors influencing the decision process, which is the case in the problem of network selection in an all-IP heterogeneous networking environment. They provide a deterministic solution that can easily be combined with fuzzy techniques and parameter estimation in cases where reliable or crisp input data are not available. The key difference amongst MADM algorithms is that they are based on different decision making philosophies. This allows the decision maker to choose the algorithm that is closest to its philosophy on decision making. In almost all the cases they are easy to understand and implement and simple enough to allow automated decision making in real time. We have avoided any discussion on a direct comparison between various algorithms discussed in the thesis on the basis of their accuracy of results since the issue has been studied in the past with the conclusion that such comparisons are not possible as they would need the use of another MADM algorithm which creates a paradoxical situation. However we have compared the MADM algorithms in terms of their suitability to solving the problem of network selection.

During our research the areas of focus related to algorithmic aspects of network selection were as follows;
• Development of a comprehensive solution for network selection using MADM algorithms.

• Modification of MADM algorithms for application to network selection when complete information about the parameters impacting network selection is available.

• Development of a new decision tool incorporating a fuzzy MADM algorithm with input parameter estimation for network selection in scenarios where complete information about the input parameters is not known.

Sometimes the problem of network selection in a heterogeneous environment has been considered synonymous to the problem of inter-technology handoffs. Although the problems of network selection and inter-technology handoffs are similar in some ways, the two have significant differences as well. For example, inter-technology handoffs have a requirement to have session/service continuity and therefore a requirement to have a low latency decision process and maintain the same IP address during a handoff. No such requirement exists for network selection at initial start up of the device, the problem that is being addressed in the thesis. An inter-technology handoff with service continuity can result in both business (e.g., higher rate of charging) and service (e.g., degradation of QoS) impacts and therefore in many scenarios this should be avoided if possible. One of the aspects of network selection as discussed in Chapter 3 can therefore be to select a network that minimizes inter-technology handoffs. It is possible to reuse aspects of the network selection solution for inter-technology handoffs where they meet the requirements such as those related to latency. However the focus of our research has been on the problem of network selection and the application of the proposed solution and algorithms to inter-technology handoffs has not been analyzed.

1.3 Main Contributions

The main contributions of the work are:
Architectural framework

The problem space related to the topic of the thesis has not been widely studied in the past and therefore is not well understood. We looked at the current state of the arts and developed requirements for a feasible architectural solution to the problem. In the process we contributed to the ongoing work\(^1\) [1] in IETF for defining the broader problem space of network discovery and selection. While considering the architectural solution in our research, we kept in mind the design constraints identified by the IETF draft.

An architectural framework for network selection in a heterogeneous wireless system has been proposed by us in Chapter 2. It is based on the network assisting the terminal in the decision about network selection. The framework has been shown to work under different deployment scenarios. The proposed solution minimizes usage of the wireless links and would work with currently deployed infrastructure. It is scalable, flexible and supports roaming. The architecture proposed by us allows for information exchange between the terminal and the network both at layer 3 and at layer 2. For information exchange at layer 3, it leverages the concept of Information Services (IS) from IEEE 802.21. IS allows for a client in the terminal and an information server with a database in the network, a notion to which we also contributed\(^2\) [23] in

---

\(^1\) The author of this thesis was one of several co-authors to the referred industry standards submission and therefore does not claim authorship of the entire referred document or the ownership of all the ideas presented in it. See reference for the complete list of authors.

\(^2\) The author of this thesis was one of several co-authors to the referred industry standards submission and therefore does not claim authorship of the entire referred document or the ownership of all the ideas presented in it. See reference for the complete list of authors.
IEEE 802.21. Here we extended this concept by defining specific service and network information related nodes. We then tied these information nodes to the current wireless infrastructure by proposing to enhance the functionality of existing policy decision nodes to include network selection decision logic as well. For information exchange at layer 2, our architecture uses as an example the work done in IEEE 802.11u for network selection to which we also contributed\(^1\) [24]. We have shown how such layer 2 solution can be integrated in our proposed architecture involving an enhanced policy decision node and supporting information nodes.

**Comprehensive MADM based network selection decision process**

A decision process for a network-assisted network selection mechanism that combines non-compensatory and compensatory MADM algorithms is proposed. The two step mechanism is more comprehensive than work published earlier. Scenarios with different services and subscribed QoS profiles have been used to demonstrate how the proposed mechanism would work. Application of the proposed mechanism to networks with multi-level Service Level Agreements (SLAs) has been described. Possible approaches to network selection when a service is already in use have been discussed and solutions are proposed.

\(^1\) The author of this thesis was one of several co-authors to the referred industry standards submission and therefore does not claim authorship of the entire referred document or the ownership of all the ideas presented in it. See reference for the complete list of authors.
Improvements to MADM algorithms for application to network selection when complete information about the factors impacting network selection is available

We have proposed an improvement to the standard Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) algorithm when it is applied to network selection by adopting an iterative approach. To the best of our knowledge this is the first such improvement of the algorithm for dealing with ranking abnormalities when complete ranking is not required.

We have identified the need for support of non-monotonic utility for attributes used in network selection decision making. We have shown that many of the commonly used MADM algorithms in their standard form are not well suited because of assumptions about monotonically increasing or decreasing utilities of the attributes. We also show that Grey Relational Analysis (GRA) is better suited for achieving this type of optimization objective. A two step process has been proposed for tuning the algorithm. To the best of our knowledge, this is the first known use of non-monotonic utility in multi attribute network selection.

We have modified Elimination and Choice Translating Priority (ELECTRE), another well known MADM algorithm so that it can be now used for network selection with attributes exhibiting a non-monotonic utility. This modification makes the algorithm more adept to application in network selection as it expands its applicability to a wider range of optimization objectives. To the best of our knowledge this is the first known application of ELECTRE to network selection and first known modification of ELECTRE for handling non-monotonic attribute utilities.

Our contributions in terms of modification to TOPSIS and ELECTRE and the usage of GRA with non-monotonic utility are not specific to the problem of network selection but these ideas can be carried over to other problem spaces with similar requirements.

Application of data prediction and fuzzy techniques in MADM based network selection for scenarios when some of the information to be used in network selection is missing

We have developed a decision strategy for network selection in the absence of precise input information. The decision process proposed by us describes how fuzzy techniques along with data prediction can be useful in network selection only in certain cases. We have developed a new function, Confidence Level, as an additional decision support function and have used it along with the sensitivity of the service/subscription to the missing/imprecise information. We have proposed that these two functions together can be applied to ascertain the usefulness of estimating the data.
To the best of our knowledge this is the first application of functions such as sensitivity and Confidence Level to the problem of network selection with imprecise information.

The algorithms presented in the thesis can be applied irrespective of the adoption of the architectural framework proposed in Chapter 2. In other words the algorithmic approaches, with some adjustments would be applicable even when the decision making entity applying the algorithms resides in the terminal. In that scenario however all the information has to be made available to the terminal.

1.4 Organization of Thesis

Figure 1.1 Organization of thesis

Figure 1.1 illustrates the overall organization of the thesis. The remainder of this thesis is organized as follows. In Chapter 2, we propose a framework for network selection in a heterogeneous wireless access system.

After laying down the architectural framework, the remaining chapters focus on the algorithmic aspects with application of multi attribute decision making algorithms to the problem of network selection. Chapter 3 explores the application of a combination of compensatory and non-compensatory MADM algorithms to provide a comprehensive solution to the problem of decision making for network selection.
Chapter 4 provides enhancements to the MADM algorithm for application to network selection when input information to these algorithms is reliable and available. It suggests improvements to the standard TOPSIS algorithm when it is applied to network selection by adopting an iterative approach. Scenarios with non-monotonic utility for QoS related attributes are discussed and it is shown that amongst the algorithms considered, GRA is best suited for non-monotonic utility. The chapter also describes the use of ELECTRE to network selection and suggests modifications to it so as to make it better suited for use with non-monotonic attribute utility.

Chapter 5 takes into consideration scenarios when input to the MADM algorithms is not completely available or trustworthy. It proposes the use of data prediction and fuzzy MADM techniques and then analyses the use of such techniques with GRA for the purpose of network selection. It also proposes new decision support tools as part of a comprehensive decision strategy for such scenarios.

Chapter 6 concludes the thesis by summarizing the work done and identifying possible future research work in the area.
1.5 References


2 Architectural Framework for Network Selection in Heterogeneous Wireless Systems

2.1 Introduction

The process of network selection in today's wireless systems is quite simple. In the case of a cellular system, e.g., GSM, network selection can involve a scan by the terminal for the PLMN IDs of available networks followed by a selection of one of them based on already provisioned information in the terminal about the preferred and forbidden PLMN IDs. There are several reasons that such a simple selection process works well. The search and hence the scanning for the best network is limited to only one type of access technology. The roaming agreements are limited to operators of one type of access technology. Therefore support for only one type of authentication credentials is needed and billing infrastructure across the roaming partners is also aligned. The most common service to be accessed is circuit-switched voice. Unlike packet-switched all-IP networks, the issues related to QoS do not arise in circuit-switched networks. These simplifying assumptions made for network selection in today's cellular systems would not be possible when delivering services over a heterogeneous mix of all-IP wireless access technologies for nomadic and mobile users.

1 Parts of this chapter have been published in the following
As an example, consider the situation shown in Figure 2.1, where a user is in a convention center served by multiple WLANs and WWANs. The user terminal supports both WLAN and WWAN access, and the user decides to access a streaming video service provided by his HN to watch the news. Currently, there is no automated way for the user terminal to intelligently select the most appropriate network from a heterogeneous mix of wireless access technologies for accessing the desired service. Such a network selection mechanism is essential to providing optimized service delivery along with a good service experience.

The remaining part of this section highlights factors that can have an impact in network selection for a heterogeneous wireless IP networking environment. The following sections describe related work, and present the new architecture along with deployment scenarios for the architectural nodes. The initial deployment scenarios do not include a network-based decision making entity that can coordinate the information from different information resources and hence require for the terminal to directly obtain the processed information separately from different information sources such as those related to services and network conditions. In the final policy-based network selection example we describe how the role of an existing cellular network entity can be enhanced to make network selection decisions. This model further improves the decision process by centralizing it and decreasing the communication overhead with the terminal. In the following chapters, we assume the usage of the architectural model described in the policy-based network selection scenario.
2.2 Factors Impacting Network Selection

2.2.1 Access Network Characteristics

Network capabilities (e.g., total bandwidth) and current network conditions (e.g., network utilization and traffic load) can play a significant role in the network selection process. Different wireless access technologies can have different capabilities. Even for the same access technology, different capabilities can exist. For example, in the case of IEEE 802.11 WLANs, 802.11a has 54 Mbps capacity while 802.11b has 11 Mbps capacity and they also operate at different frequencies and therefore have different requirements on the terminal.

2.2.2 Service Selection

Unlike circuit-switched services such as traditional voice services, services delivered over an all-IP network, such as voice over IP (VoIP), will be strongly influenced by the network conditions and capabilities. The initial selection of the wireless AN by the terminal can therefore be impacted by, e.g., the QoS requirements for the service that the user has decided to use. In addition, different wireless ANs may support different sets of well defined IP services (e.g., VoIP based on IP Multimedia Subsystem (IMS)) at the same geographic location and a user decision to use a different service at a later stage of network attachment can result in the current network becoming sub-optimal or not available. For example if the initially selected network does not have an appropriate level of QoS or cannot support voice for regulatory reasons, then the user’s decision to make a voice call may trigger a network re-selection process.

2.2.3 Roaming Support

The terminal should be able to get service over its HN or the most optimal roaming partner network. Roaming support is considered essential for ubiquitous service delivery as no single network operator can afford the infrastructure to provide global coverage of its customers. A key benefit of roaming support is the ability to share revenue among the partner networks while providing a better coverage to customers.

2.3 Background and Related Work

While there has been work done on individual aspects of heterogeneous wireless networks such as the network selection algorithms, little has been published on system level issues that would call for an architectural solution for interworking of heterogeneous wireless networks operating
in autonomous domains. However, [1] does provide such details at the systems level by laying out the requirements for seamlessly accessing a communication system that has a heterogeneous mix of wireless access technologies. One of the key requirements the paper defines is the ability to discover all the radio networks available in an area. To meet this requirement, the paper describes a proof of concept system that uses BAS to identify all the radio technologies. BAS runs over a basic AN, which can employ either a current radio technology or a new one. The approach taken in the paper leads to significant shortcomings in the architecture. For example, the paper does not take into account the fact that networks are typically operated and managed by autonomous business entities, and hence fails to consider the need for supporting multiple authentication mechanisms over roaming interfaces with users having different security/authentication credentials. The architecture also does not include the concept of a user having a HN where the user is subscribed to a set of well defined home based services with different QoS requirements. These are services that the subscriber may want to access from different wireless access technologies owned and operated by autonomous business entities.

2.3.1 Related Recent Standards Activities

The protocols and mechanisms being developed in several industry forums can possibly be used in the network selection architecture proposed here. The IEEE 802.21 [2] group has been working on the problem of inter-technology handoffs. New work has been proposed in the IETF Mobility for IP: Performance, Signaling and Handoff Optimization (MIPSHOP) [3] working group to carry out network layer protocol work to transport IEEE 802.21 related information. The IETF Protocol for carrying Authentication for Network Access (PANA) [4] group has been working for sometime now to develop a network-layer authentication protocol that can support various authentication methods, dynamic service provider selection, and roaming clients. The Internet Research Task Force (IRTF) Mobility Optimizations (MOBOPTS) [5] group is also looking into issues related to optimization of mobility across heterogeneous networks. Ongoing work in the IETF Next Steps in Signaling (NSIS) [6] group is performing some of the QoS mapping related work across access technologies. 3GPP [7] is currently working on an evolution of an architecture that could support multiple radio technologies. IETF draft [8] provides a good description of several aspects of the problem space that need resolution and provides the requirements for an acceptable solution. The IEEE 802.11u [9] task group is currently working on solutions that could amend the IEEE specifications so that IEEE 802.11 based networks could
better interwork with other systems. Below we describe the relationships of some of the standards related activities to our proposed architecture.

An important aspect of our architectural framework is that it requires the terminal to initially obtain (layer 2 or layer 3) network connectivity, with or without authentication, in order to query policy related components. This initial connection is not used for accessing a subscribed service, but to exchange information related to the network selection process. The network used for this search may not be the most appropriate network in terms of network capabilities and the user services supported. This query can be viewed as similar to using a directory assistance service.

A similar approach is currently under discussion in IEEE 802.21, the Media Independent Handoff (MIH) working group, which is developing standards for assisting handoffs between heterogeneous networks. The group is defining three types of information that can be shared across different networks: Event Services (ES), Command Services (CS), and Information Service (IS). IS provides less time critical information, ES provides information needed in real-time for handover decisions across different network types while CS provides commands for the actual handover. The standards allow all three types of information to be delivered through either link-layer specific solutions or through a “layer 3 or above” protocol. As part of our work we also contributed in IEEE 802.21 to the notion of an IS with a terminal-based client being able to query at layer 3 a network-based information server connected to a database [20]. We have used this concept in our framework for services and network related information nodes.

The IETF MIPSHOP working group [3] is defining the delivery of MIH information at layer 3, which enables MIH services even in the absence of link-layer support, while the IEEE 802.11u working group is developing layer 2 standards for interworking of WLANs with external networks such as cellular systems, including the ability of a terminal to discover and select a network before associating to a WLAN AP.

The IEEE 802.11u effort is focused on the support of MIH in WLANs. The current draft standard proposes the use of a Generic Advertisement Service (GAS) functionality to query network related information. It specifies the transport mechanisms for APs to advertise services for network selection or other purposes to WLAN clients in the un-associated state. 802.11u GAS capability is included in the 802.11 beacon and probe response frames. Although GAS transport is transparent to the Advertising Protocol used for queries and query responses and is flexible enough to support multiple protocols, at this time only 802.21 is developing such a
protocol, but for the purpose of supporting network selection for handover decisions rather than initial service access. As 802.11u has a requirement to support 802.21 based mechanism for network selection in WLANs, it is of interest to leverage this capability for network selection at initial service access. Following the current 802.11u draft proposal, terminals obtain the GAS information, particularly the identity of the supported Advertising Protocol, from beacons broadcast by 802.11u compliant APs or by exchanging probe response messages with such APs. Using the specified Advertising Protocol, the terminal sends a GAS-Initial-Request with a specific query to the GAS-aware AP, which then forwards the request to a back-end GAS policy server to retrieve the query results. The response message from the policy server is returned to the terminal via the AP using the Advertising Protocol. We were one of the contributors to the proposal for this layer 2 based network selection proposal in IEEE 802.11u that uses GAS [21]. In this Chapter we demonstrate how the mechanism as proposed in IEEE 802.11u can be used within our framework.

2.4 Proposed Architecture

The problem of network selection in a heterogeneous wireless networking environment can be solved in two ways.

2.4.1 Terminal Based Approach

In this approach, the terminal itself first discovers all the available ANs. It then collects from the networks individually the information necessary for making a decision and chooses the most appropriate AN after analyzing the collected information. This approach is used today in homogeneous wireless networks quite successfully. In such systems, there are only a limited set of frequency ranges for the terminal to scan, all the systems have typically similar capabilities, a single authentication mechanism is uniformly supported by everyone (e.g., SIM for GSM systems), and the roaming agreements are limited to operators using the same technology. In the case of GSM networks, the network selection process can simply be a table lookup comparing a list of preferred PLMN IDs with the broadcast PLMN IDs received by the terminal. There are however significant problems with this approach when applied to heterogeneous all-IP networks where the number of variables in the network selection process are significantly higher. A multimode terminal will typically not have a priori knowledge of all the information related to a heterogeneous mix of roaming partners, e.g., frequencies of operation (scanning may require turning on multiple radios), network identities, authentication mechanisms (it may require
several retries to find the supported mechanisms), etc. Unlike today's mostly circuit-switched service networks used for voice calls, a simple availability of an IP network for use does not indicate how good or bad network conditions are on the available network or its capabilities to support a variety of IP services with different QoS requirements. The process of network discovery in this case would therefore likely be an unpredictable one with several possible retries and failures.

For most wireless systems, the wireless link is the most expensive and bandwidth limited part of the transport and therefore use of this link has to be efficient. The mechanism whereby the terminal individually queries each of the available ANs can place a significant extra overhead on the wireless links. In addition, for reasons of security and lack of air link resources, the network operators would be reluctant to provide information about the network conditions directly to an individual user, especially if he is a roamer and has not yet authenticated with the network. Even if the network operators do allow the sharing of such network related information, current MAC/PHY layers of deployed access technologies do not have the ability to provide such information to a terminal before authentication. Any suggested changes to the MAC/PHY layers would take several years to standardize and additional time would be needed for product roll out and deployment. Such solutions may also not be backward compatible with the deployed infrastructure and terminals.

2.4.2 Network Assisted Approach

In a network-assisted approach, the network assists the terminal in the decision process by performing data collection and analysis. Along with its location, the terminal can also provide any other information that could be considered by the network in the analysis, e.g., service/QoS being requested, scans of the PLMN IDs, SSIDs, etc. The network only assists the terminal in the decision process and the final selection is still done by the terminal (or the user). The final selection of an AN from a preferred list provided by the initial network can be made by the terminal-based on additional information such as the signal strength from different networks. In this approach, the terminal would try to get network connectivity (at layer 2 or layer 3) with whichever AN it can (with or without authentication). This initial network access can be compared to getting directory assistance in today's world of circuit-switched voice calls. The initial AN may not be the optimal network for the requested service and can be changed once the terminal has received sufficient information from the network entities to make a decision. While
the deployed ANs vary in nature, e.g., WLAN, UMTS, WiMAX, a unifying aspect of different access technologies being deployed is that they support IP transport. This makes the network-assisted approach for network selection using layer 3, i.e., IP layer entities, feasible. The network-assisted approach can be much more bandwidth efficient for the wireless link as the terminal would not have to query individual ANs for information related to network selection.

Consider a scenario of $n$ alternative ANs with network selection related query / response messages to each of these networks taking an average of $m$ bytes. The total wireless bandwidth consumed in this case will be $n \times m$ bytes for the terminal-based approach. For the case of network-assisted approach, the terminal will be able to access similar information in approximately $p \times m$ bytes where $1 \leq p \leq n$. The value of $p$ will depend upon different deployment strategies. The static or provisioned component of information (mainly capabilities of the network) in this approach could be stored centrally within an operator network, for dissemination based on policies and agreements. The frequency at which this information is exchanged with other domains would depend on the type of information and how often it is refreshed. Both pull and push models for information exchange could be supported. The main drawback of this approach is the relative delay introduced in the network selection process when compared to the terminal-based approach. Since the collection and dissemination of dynamic network conditions through this architecture uses IP, i.e., layer 3, it may not be fast or frequent enough to support handover decisions. However, the latency requirements for initial network selection are not as strict as for handoff and therefore this mechanism would work adequately for the purposes of network selection at the time of initial service access for both nomadic and mobile type of users. Table 2-1 compares the two possible solutions. Based on this comparison, the network-assisted approach is preferred. The rest of the chapter discusses architectural solution based on this approach.

In order to better understand the proposed network-assisted selection solution, the decision process is shown in Figure 2.2 at a high level so that it is applicable to many of the possible deployment strategies. The term network selection node has been used in Figure 2.2 as a general term to accommodate various possible deployments of the logical nodes described later in this section. The blocks shown in Figure 2.2 do not represent a logical or physical network entity but the actions performed during various stages of the selection process. The flowchart in the figure assumes that the terminal is capable of exchanging information with the network via layer 2 or 3 mechanisms.
Table 2-1 Comparison of terminal-based and network-assisted mechanisms

<table>
<thead>
<tr>
<th>Backwards compatible and work with current infrastructure</th>
<th>Terminal Based</th>
<th>Network Assisted</th>
</tr>
</thead>
<tbody>
<tr>
<td>This mechanism works better with a layer 2 solution for information exchange and therefore is less likely to work with current deployments because of required changes to MAC/PHY layers of access technologies.</td>
<td></td>
<td>This mechanism can work well both at layer 3 or layer 2 but when used at layer 3 it can work with current systems because the IP layer is common to all broadband wireless systems.</td>
</tr>
</tbody>
</table>

| Flexible deployment and support for different business models | Required changes for this mechanism will likely happen at the edges of the networks. | IP nodes can be placed possibly in different networks depending upon the deployment and business strategy. |

| Scalable based on network and computational resources | This approach consumes more bandwidth resources over the wireless links along with more computational resources of the terminal. | Most of the data collection and processing is done within the wireline network and therefore wireless bandwidth resources and terminal computational resources are not impacted significantly. |

| Support for roaming | This mechanism may not work in roaming as the networks may refuse to share information prior to having a credible request from a known entity. | This mechanism will work in roaming as the information is shared with other network operators or trusted entities over the IP network. |

| Latency in decision making | For this mechanism, latency per network query is expected to be less as only MAC/PHY layers are involved but it requires an exhaustive scan for all networks. | For this mechanism, while using layer 3, latency is expected to be more as IP layer communication is involved with a backend information server. |

Figure 2.3 shows a high level architecture having several autonomous networks connected to heterogeneous wireless access systems. The figure shows the relationships between the ANs, their roaming partners and the HN of a user accessing the network. Figure 2.4 provides further details about the logical nodes within the architecture that are relevant to the network selection process. More specifically, the proposed architecture introduces two new nodes, Service Announcement Node (SAN) and Data Collection Node (DCN). It also adds some new functionality to the user equipment (UE) and the AAA node. The new nodes represent logical functionality that will reside in the network and they are described later in this section.
Network selection request is received from the terminal. It includes information such as terminal location, terminal capabilities and the requested service.

Network selection nodes retrieve provisioned / static information about candidate access networks and services that they support.

Network selection nodes refine list of candidate networks by comparing information received from the terminal such as the access network types, user credentials supported, services requested etc. with similar information about access networks.

Network selection nodes retrieve information on dynamic parameters for remaining candidate networks (e.g. network congestion) that can influence network selection.

Network selection nodes further refine the list of candidate networks using dynamic parameters as a selection criteria.

Network selection nodes communicate the list of the networks that fulfill the minimum selection criteria to the terminal or pick only the best network and communicates its identity. Other pertinent information such as access technology, SSID (in the case of WLAN), authentication mechanism etc is also included.

Figure 2.2 Decision process in network selection using proposed network-assisted mechanism

Figure 2.3 A heterogeneous wireless system with access networks using different technologies
The system as represented in the two figures does not include all the logical functional nodes that would be present in a heterogeneous networking environment. Only those nodes that are related to the network selection process have been presented. Other functional nodes, e.g., mobility management node, are outside the scope of this discussion.

2.4.3 Using DCN to Collect and Provide Access Network Characteristics

The processes of network monitoring, and data collection, storage and presentation are performed by the functional entity DCN as shown in Figure 2.4. DCN functionality will primarily have two sub-functions. Firstly, it monitors and collects information on network conditions using a standard set of parameters so that these parameters have the same interpretation across different autonomous networks using different access technologies. This sub-function is named the Information Collection Function (ICF). Secondly, it stores and exchanges this information across autonomous domains in a standard manner so that the information can be exchanged easily across the boundaries of autonomous systems. This function is named the Information Provider Function (IPF). Data collection will be done in the ANs but the collected information can be stored anywhere in the network. The ANs such as WLAN hotspots can have SLAs with partner networks that could allow certain QoS guarantees (e.g., in terms of latency, jitter, and packet drop rate). While the AN would mainly provide information about the conditions and capabilities of its own network, it can also include
information from its SLAs with partner networks or the network conditions it is experiencing with its partner networks.

A good basis for selecting appropriate QoS related attributes is provided in [10] and [11]. The attributes discussed in [10] and [11] can be used to describe network capabilities and current network conditions across a variety of AN types. The following is a subset of the characteristics described in [10] that are relevant to this discussion.

- Bandwidth Characteristics
- Delay Characteristics
- Loss Characteristics
- Availability Characteristics
- Packet Re-Ordering Characteristics

The discussion on a possible list of parameters used to measure these characteristics can also be found in [10]. In [12], methods to collect and distribute such network related information have been discussed.

### 2.4.4 Using SAN for Service Selection

Although the concept of service announcement is not new, when requested, the SAN in the proposed architecture provides information about well defined user subscribed services that are available to the requesting user via a particular AN based on not only the operational requirements such as network capabilities but also the business requirements such as legal and contractual constraints. In some scenarios there can be a good mapping between network capabilities and the types of services that the network supports so that the supported services can indirectly be derived from network capabilities information. However, other factors such as business models can also influence the services that are offered by a network in a geographic location. In some cases due to regulatory constraints the network may explicitly indicate no support for a service even if the network capabilities could support the service from a technology perspective. For example, based on the business model or regulatory issues (e.g., CALEA, E911), the AN may choose to be either a conduit for data or it may decide to provide services such as streaming media and VoIP directly or via its partners (e.g., content aggregators) as well. The SANs would be used to announce the presence of such services. Depending upon the
deployment scenarios as documented, the SANs being queried by the user terminal can reside in
the AN, its partner networks, or the user HN. Existing IETF protocols such as LDAP or SLP
could possibly be reused for acquiring such service related information.

2.4.5 Using Enhanced AAA for Roaming Support

The functionality of AAA nodes in the proposed architecture is similar to the one that exists
today in data networks but enhanced to handle heterogeneous systems. In a heterogeneous
wireless environment, different user authentication credential types and mechanisms may be
used. For such systems to interwork, the user terminals as well as wireless ANs must have a
common or interoperable way of providing to each other information such as their identities,
their HNs, their capabilities and their roaming agreements, etc. Network layer mechanisms can
be used to provide such interoperability. An example of providing user identity and HN identity
at the network layer can be the use of NAIs defined in IETF RFC 2486bis, where the realm part
of the NAI can provide the name of the HN operator. A decorated NAI can also include
information on how to route the request via multiple roaming partner networks. Once the AN
operator has the information on HN and roaming partners, it can decide if it has any agreements
to authenticate and charge the user for its usage of the network. The HN can also communicate
service and subscription related information in its AAA messages. In order to get ubiquitous
coverage by roaming with a large number of roaming partners, it may become necessary for a
multimode terminal user to carry more than one type of authentication credentials for the same
HN (e.g., SIM, user ID/password). Protocols such as PANA [4] with DIAMETER/RADIUS in
the backend can be used in the enhanced AAA nodes proposed to provide such support. [8] has
discussed in detail issues associated with AAA routing. Since any proposed solutions have to
apply to a wide range of AAA messages, any enhancements to resolve them should not restrict
the introduction of new AAA or access network functionalities.

By using the new AAA functionality along with available service and transport level information
via bearer path nodes (e.g., routers, application servers), more complex charging relationships
can be supported. For example, it can be possible to perform differential charging or content
based charging by correlating IP flow with the content downloaded and then charging the
customer for the content only.
2.4.6 Information Collection and Exchange

For network selection, the system collects and provides two types of network information: static network capabilities, and dynamic network conditions. It is necessary to get reliable and updated values for both types of information before the process of network selection.

The collected information to be used in the decision process can be stored anywhere in the network. The ANs will have SLAs with partner networks that provide certain QoS guarantees (e.g., latency, jitter, and packet drop rate). While each AN would mainly provide information about its own conditions and capabilities, it can also include information from its SLAs with partner networks, or on the network conditions it is experiencing with its partner networks. Figure 2.5 shows, at a high level, how information to be used in policy-based network selection is collected within an autonomous domain and then shared across the domain boundaries based on operator policies and roaming agreements. In [12], similar mechanisms to collect and distribute resource information have been discussed, although for a different usage. However the mechanism is well suited for network related information collection for network selection as well.

The process can be broken down into the following steps:

- Monitor resources within the AN. The raw information may be processed for data reduction before collection by central nodes within the domain.
- Based on the information received from ANs within the domain, model and forecast/predict availability of these resources.
- Disseminate the information to roaming partners using standard parameters, based on policies and roaming agreements.

Figure 2.5 also shows that static or provisioned information about the services and ANs is stored centrally within the autonomous domain, for dissemination based on policies and agreements. The frequency at which this information is exchanged with other domains would depend on the type of information and how often it is refreshed. Both pull and push models for information exchange could be supported. Figure 2.6 illustrates the fact that while the QoS attributes used for functions such as network selection, admission control, handoffs, and packet scheduling are likely to be similar, because of the differences in functional requirements the update rates and hence the levels of accuracy will be different.
Figure 2.6 QoS attributes used for network selection, admission control, handoffs, packet scheduling can be the same but will have different update rate and the level of accuracy.

The collection and dissemination of dynamic network conditions through this architecture can occur at layer 3, and may not be fast or frequent enough to support real-time handoff decisions. However, this information could sufficiently reflect the dynamics of the network to facilitate network selection at the time of initial service access. For example, if average utilizations of two
ANs, A and B, with similar capabilities and bandwidth are 90% and 30%, respectively, then the terminal should choose network B to access service.

2.5 Deployment Scenarios

In order to discuss deployment scenarios for automated network selection, we again consider the example briefly described in the introduction where a user is in a public place served by multiple WLANs and WWANs. The terminal supports both WLAN and WWAN access and the user decides to use his HN based video streaming service to watch the news. Currently, there is no automated method for the terminal to intelligently select the most appropriate network from a heterogeneous mix of wireless access technologies for the desired service. Without an appropriate architectural solution, the user may have to manually go through each individual AN and select the one he considers acceptable. The process can be quite inefficient and can result in excessive network traffic and subsequent delays in the process of network selection. The information available from individual ANs for network selection purposes may also be non-uniform. This could result in the selection of a suboptimal network, e.g., in terms of network resources or cost. In many cases the user may not be able to find out beforehand if certain services are supported by the AN. Because of the non-uniform information availability for network selection, the process may not be consistent and may not work all the time under roaming conditions.

Using the proposed architecture, different deployment for the SAN and DCN nodes can occur. The following sub-sections describe the more common deployment scenarios and the decision processes involved in them.

2.5.1 DCN (IPF) and SAN residing in the AN

Figure 2.7 shows this scenario. The user terminal first tries to get connectivity with an AN. In secure ANs this starts with the process of authentication. The decision to initially select one of the ANs for network connectivity can be a user terminal behavior based on possible information available to the user terminal via, e.g., provisioning, prior experience with the networks or random. Once the user terminal gets connectivity to an AN (e.g., it is assigned an IP address), it requests and gets the information about services supported by the AN from the SAN residing in the AN. For example, the SAN can inform the user that the network supports VoIP using IMS. The user terminal also requests and gets information about the current network conditions from...
the DCN located within the AN. This allows the user terminal to make a decision if it should proceed with initiation of the service, e.g., a HN based streaming media service. The user terminal then proceeds with initiating access to the HN based service. The information provided by the SAN and DCN in this case may be limited to the AN in which it resides. The user terminal can also proceed with authenticating with other ANs in order to get information from their DCN and SAN nodes before making a final decision on the AN to use for the service.

Figure 2.7 DCN (IPF) and SAN residing in the AN

2.5.2 DCN (IPF) and SAN residing in local information provider network

Figure 2.8 shows the deployment scenario. Once the user terminal gets connected to the AN (e.g., it is assigned an IP address), it gets the information about the services supported by the AN from the SAN residing in the network of the local information server.

A possible variation to this scenario can be that the local information provider may have deployed an AN (such as a WLAN) that provides local connectivity to anyone without pre-authentication to provide access to the SAN. The local information server can be a free service provided by the owner of the premises or a local communication broker where the user terminal is trying to get the service. The information service provider may not own any of the ANs. For example, a SAN node operated by the owner of a convention center and accessible via the local AN operators, can inform the user that the AN operators X, Y and Z support VoIP using IMS in
the premises. If the regulations require support for emergency calls, this would imply support for location information by the ANs. The SAN can also provide further information regarding, e.g., the frequency bands of operation and the pricing for these networks. The user terminal also gets information from the local information server about the network capabilities and current network conditions in multiple ANs operating in the current location. Additional information such as the type of credentials and authentication mechanisms used by the networks may also be provided. This allows the user terminal to make a decision if it should proceed with initiation of the service, or if it should switch to a different AN before service initiation.

![Figure 2.8 DCN (IPF) and SAN residing in local information provider’s network](image)

### 2.5.3 DCN (IPF) and SAN residing in the HN

This type of deployment, shown in Figure 2.9, allows the HN to manage the user experience by managing network selection related information that is provided to the user terminal. The user terminal or the AAA proxy server of the AN being used is able to provide the user’s geographic location during the authentication phase or afterwards. After getting connectivity to the network via layer 2 or layer 3 mechanisms, the terminal accesses the information entities located in its home network. The HN based DCN (IPF) and SAN, based on the roaming relationship, are able to collect information about possible local ANs on which the user can roam, their network capabilities / conditions, and the services that they offer. Based on the information received from
its HN, the terminal can make a decision to change the AN to the one better suited for the service to be used.

In all the scenarios described above, the ability of the terminal to get to the appropriate information providing entities in the network can be based on a Domain Name Server (DNS) lookup that would allow a terminal to contact, e.g., “services.home_operator_name.com” for service related information in the home network and “service.local.com” for services information in the local AN. Also, to handle services that can trigger a new session from the network side (e.g., a terminating VoIP call) the user/terminal should indicate the intention to be able to accept such a service, at the time of network selection request. This can be part of a pre-configured user profile that can include information such as his mobility (e.g., nomadic) and preferred services (e.g., voice) that can then automatically be used at the time of network selection request.

In this section we have addressed the problem of automated network selection in an environment where heterogeneous all-IP wireless ANs are available to a user terminal. Limitations with the current state of the art have been highlighted. Two approaches to solve the problem are discussed and a network-assisted network selection mechanism is proposed that defines new network layer nodes as well as new functionality for current nodes. The proposed solution provides a common architecture for automatic network selection in a heterogeneous wireless system where the user terminal requests and receives information from the network

![Figure 2.9 DCN (IPF) and SAN residing in HN](image)
before selecting an AN for service initiation. It leverages the ongoing protocol development in IETF and IEEE. Various deployment scenarios have been discussed to explain how the proposed architecture would work. The proposed architecture minimizes usage of the wireless links and would work with currently deployed infrastructure. It is scaleable, flexible and supports roaming. In the prior example, a network-based entity that can coordinate the exchange of information across different information resources (such as the SAN, DCN, etc.) has not been described and essentially each information entity is assumed to independently process the information and then interact with the terminal. In the following section we analyze a policy-based network selection architecture that leverages currently proposed elements of a 3GPP system and provides for a network entity that can coordinate across different information resources to handle terminal request for network selection.

2.6 Policy-Based Network Selection

Future mobile devices will be equipped with a multitude of wireless access technologies that enable ubiquitous service access via alternate wireless ANs supporting an all-IP data transport. This common data transport capability allows the transport plane to be decoupled from the service plane in the next generation wireless ANs, facilitating delivery of converged services over a variety of access technologies. Services such as those based on the IMS are agnostic of the type of the underlying AN so long as it supports certain capabilities, e.g., a certain level of QoS. This scenario presents a complex but interesting set of management problems, with the selection of an optimal AN for service access being a key one. A network selection mechanism that uses operator policies to provide ubiquitous services using the best among a mix of coexisting heterogeneous wireless networks for service delivery at any location does not currently exist. Today's mobile devices mostly support only single mode wireless access and employ simple network selection policies that would be insufficient to meet the much more complex network selection scenarios encountered by future multimode devices. For example, a GSM cellular device would scan for unique network identifiers broadcast over alternate ANs. These network identifiers do not provide any information about current network conditions, capabilities or the services supported, as such information is typically not needed in today's homogeneous ANs in which the primary service, i.e., voice, is supported universally using circuit-switched airlinks with tightly controlled QoS. In addition roaming agreements amongst operators are mostly applied across one access technology and therefore the need for supporting multiple credential and authentication types does not arise either. In the example cited the network selection
decision is made entirely by the terminal-based on scanned information about network identifiers. In order to develop an intelligent network selection policy that is better suited to a heterogeneous all-IP wireless access environment, more information about the alternate ANs would be needed. Such information can be dynamic and may be more readily available within the network nodes rather than pre-provisioned within the terminals.

In prior related work 3GPP has documented a network selection solution [13] for WLANs interworking with cellular systems. Broader network selection issues for cellular systems [14] are also being discussed within 3GPP.

In this section we combine policy related components defined in recent standards activities with the novel network selection components introduced in the previous sections to enable automated network selection. The original solution introduced here allows the UE to intelligently and automatically select from several available ANs, through a policy-based decision mechanism using multiple attributes such as the service to be accessed, the ANs' QoS capabilities and the current network conditions. We solve the problem by proposing a policy-based network selection mechanism whereby the network assists the terminal in making an intelligent choice about the AN for service delivery, and identifying the information components needed.

The factors to be considered in a policy-based network selection include service aspects, access network characteristics, costs and roaming support. They have been described in the beginning of this chapter.

2.6.1 Adding Standards-based Policy Nodes to Network Selection Framework

Figure 2.10 depicts a heterogeneous wireless access system that can deliver services to a user over several available autonomous networks, some having roaming agreements with the user's HN while others do not. Figure 2.11 illustrates the service, policy and transport layer logical components involved in network selection in such an environment. For clarity, logical nodes that would be present in a heterogeneous networking environment but are not involved in the network selection process, e.g., those supporting mobility management functions, have been excluded from this figure. The service and network information nodes for policy-based control are new functions introduced for network selection. They have been described in earlier sections of this chapter. In addition the traditional functionality of Policy Decision Function (PDF) as defined by 3GPP [15] is enhanced to support network selection by using MADM as described in [16][17][18]. Figure 2.11 also shows the HN-based Call Session Control Function (CSCF), an
IMS service layer functionality [19] that interacts with the policy components.

The user accesses the network to use a particular service, e.g., to make an IMS-based VoIP call. Using the elements of the proposed architecture described earlier in this chapter, it is possible to make an informed decision based on the service related information, the characteristics of the AN and the roaming aspects. In some cases there can be a good mapping between network capabilities and the types of IMS service that the network can support. If so, supported services can indirectly be derived from network capabilities information. However, other factors can also influence the services that are offered by a network in a geographic location; e.g., an AN may not support voice services due to regulatory or contractual constraints such as a requirement to support emergency calls. Such policy information is exchanged by roaming partners. The HN can provide such service related information, e.g., about IMS services offered by it and its roaming partners to its subscribers based on their locations. The terminal, by providing its location, would be able to retrieve relevant service information. For interoperability purposes, service announcements should be made in a standardized manner, such as proposed by the IEEE 802.21 and IEEE 802.11u working groups.

Figure 2.10 High level architecture for a heterogeneous wireless system with user's HN providing the services
The PDF and Policy Enforcement Function (PEF) are standard functions in the 3GPP Policy Control and Charging architecture [15]. The PDF makes QoS and other policy related decisions for IP services such as those based on IMS. We extend the role of PDF to include network selection decision making based on network, service and user subscription related information as shown in Figure 2.11 using a MADM algorithm [18]. The PEF enforces the QoS and any other decisions communicated by the PDF. Policy decision and enforcement functions are described in [15] and their application to non 3GPP wireless access technologies is being considered in System Architecture Evolution (SAE) of 3GPP [7]. Policy decision and enforcement related functionalities have also been adopted by wire-line forums such as Cable-Labs and ETSI TISPAN. Thus an industry consensus similar to IMS is emerging in this area.

2.6.2 Example of Policy-based Automated Network Selection

Now we again consider the scenario where a user employing a dual-mode terminal that supports both WLAN and cellular access is in a public place served by multiple WLANs and cellular systems. In this case the user decides to access the IMS-based VoIP service at his HN at a location where two ANs are available, as shown in Figure 2.12.
Figure 2.12 High level communication with policy components residing in the HN. The PDF is enhanced to include the functionality of decision making for network selection. The decision making algorithms that can be used are described in later chapters.

As described earlier, different alternatives exist for deployment of the network selection policy nodes. Here we consider the case where they are located in the HN, a deployment considered suitable for HN-based IMS services. The terminal obtains the information needed for network selection from its HN. Such a deployment works well under roaming scenarios as the HN is able to collect information about services and ANs of its roaming partners as described earlier.

As shown in Figure 2.12, the terminal initiates an information exchange with an AN. In this example, the decision to select AN#1 for accessing the back-end policy server is based on information broadcast by AN#1 in its beacon. In other scenarios it can be also be a terminal behavior based on possible information available to the terminal via, e.g., provisioning, prior experience with the networks, or random. In this example, the MAC layer of AN#1 supports queries at layer 2. Once the terminal has identified the AP for AN#1 to be 802.11u compliant, it can make different types of queries; e.g., it can request information from the GAS-aware policy server in its HN about available ANs at its current location that support certain IMS services and have roaming relationships with its HN and have the capabilities and network conditions to
deliver the services. This type of information can allow the terminal to decide if it should proceed with service initiation. Based on the information received during queries to the HN, the terminal decides to change AN before initiating the service. Once the AN has been selected the terminal proceeds with its interaction with HN CSCF via the selected AN (AN#2 in this example) to initiate IMS services, as shown in Figure 2.13. The HN CSCF communicates QoS and resource requirements for the requested IMS service to the HN PDF, which has peer relations with the PDF in the visited network to facilitate negotiation of QoS and resource allocation. The PDF communicates the final QoS request to the PEF in the visited network. With the use of intelligent network selection in the prior step, the chances of success in getting the requested QoS are maximized.

![Figure 2.13 Signaling and bearer paths once AN#2 has been selected. The bearer path is route optimized via a local breakout](image)

The HN-based deployment of network selection policy functions allows the HN to manage the customer experience by providing the appropriate network selection information to the terminal. The terminal or the AAA proxy server of the AN being used should be able to provide the terminal location to the HN during the authentication phase or afterwards. The HN-based network and service information elements, based on existing roaming relationships, are able to suggest possible local ANs that the user can roam on, which offer suitable capabilities, dynamic
conditions, and service support. The central location of information in the HN allows for easy updates of roaming related information and coverage maps. The ability of the HN to process and filter information before delivering it to the user can enable efficient usage of network resources and reduce computational requirements for the terminal.

2.7 Conclusion

In this chapter we have developed an architectural framework to enable automated network selection for delivery of services in the next generation all-IP heterogeneous wireless access environment. Limitations with current approaches have been highlighted and a new policy-based solution that leverages ongoing standards activities has been proposed through the addition of new policy functions to the existing policy related nodes and the introduction of new function nodes for data collection and service announcement, which can be readily implemented in users' home networks. The novel solution presented in this chapter effectively provides the functionality of automated network selection where the terminal requests and receives a list of recommended access networks from the initial access network before selecting a service delivery network. Sample deployment scenarios have been discussed to explain how the proposed solution works.
2.8 References


3 Use of MADM Algorithms for Automated Network Selection in a Heterogeneous Wireless Network Environment

3.1 Introduction

The desire to increase service availability is driving the use of multimode terminals and the interconnection of different wireless access technologies that support IP transport. An essential aspect of service delivery in a heterogeneous all-IP wireless network environment is the selection of an optimal AN. Network selection in such an environment is influenced by several factors and currently there is no comprehensive solution available to solve this problem. Selection of a non-optimal network can result in undesirable effects such as higher costs or poor service experience. A system architecture for a terminal-based approach that uses network assistance for network selection has been proposed in Chapter 2 [1]. An overview of this approach is provided in the beginning of Section 3.3. Here we describe an extensive decision making process that can be used by the network to provide candidate networks for service delivery to the terminal.

The proposed mechanism is based on a unique decision process that uses non-compensatory and compensatory MADM algorithms jointly on the network side to assist the terminal to select the top candidate network(s). Along with disjunctive and conjunctive type of non-compensatory MADMs, we describe the use of TOPSIS, a compensatory MADM algorithm. It identifies various factors that would influence the selection of an optimal network and therefore should be used as

---

1 Parts of this chapter have been published in the following
inputs to the MADM based decision process. It explores different aspects of the problem space and proposes solutions based on MADM mechanism. The steps involved in the use of non-compensatory MADM algorithm are simple. They remove network alternatives from the candidate list that are not suited for the scenario. This process is completed before a compensatory MADM algorithm can be used to provide network ranking. The compensatory MADM part of the proposed mechanism is more sophisticated with tunable parameters. It has therefore been validated here by applying it to different use scenarios; e.g., scenarios with different service and subscribed QoS profiles have been shown to observe how the proposed mechanism would work. Application of the proposed mechanism to networks with multiple levels of SLAs or multiple classes of QoS is described. Possible approaches to network selection when another service is already in use by the terminal are presented within the proposed framework.

3.2 Background and Related Work

The problem of network selection across heterogeneous wireless networks has recently received much attention because of a drive for converged communication systems. In this context, [2] has proposed the combined application of two mathematical techniques in an algorithm for network selection between UMTS networks and WLANs, where the analytic hierarchy process and Grey system theory are used for evaluating the user's preferences and service requirements, and for combining the priority settings of the QoS attributes with the performances of the network alternatives to make network selection decision. A dynamic system to select the network for service delivery based on current market conditions such as QoS and cost attributes has been described in [3]. Using the proposed framework, the user can select the delivery network per call. Network selection based on resource allocation strategy for the most efficient resource utilization in a heterogeneous networking environment has been proposed in [4]. A fuzzy logic based multiple-criteria decision-making system to perform access network selection in a heterogeneous networking environment is described in [5]. A simple policy-enabled handoff system across heterogeneous wireless networks is presented in [6], which allows users to express policies on what is the best wireless system at any moment, and make trade-offs among network characteristics and dynamics such as cost, performance and power consumption. The mechanisms described in [2]-[6] have significant shortcomings. The factors considered in the decision processes are insufficient; e.g., information about access types supported, authentication types supported, roaming partners supported are not considered in the decision processes described in these papers. They also do not provide or refer to a viable architectural framework in which the
selection mechanism can work. The use case scenarios described are limited and not realistic from the perspective of the deployment or business models used in the industry. As a result, they do not provide a complete and deployable solution to the problem.

The decision about network selection in a heterogeneous wireless environment is dependent on several factors. The network selection problem can be solved using MADM algorithms. MADM has been an active area of research since 1970s. Because of their deterministic nature and easy implementation, MADM algorithms have found applications in a wide variety of problems, from social sciences to operations research. They can be used in combination with fuzzy logic where input attributes values are not clearly defined. Another interesting use of these algorithms involved game theory where the players manipulate input values to influence decision making by MADM algorithms. Amongst the most widely discussed MADM mechanisms are ELECTRE [7] (Elimination and Choice Translating Priority), TOPSIS [8] (Technique for Order Preference by Similarity to Ideal Solution), AHP [8], [9] (Analytic Hierarchy Process), GRA [10] (Grey Relational Analysis), WSM [8] (Weighed Sum Model) and WPM [8] (Weighed Product Model).

The mechanism described in [2] has also utilized an MADM approach for network selection but with only a limited number of parameters and without an architectural framework. Factors such as the roaming agreements of the user’s HN, the authentication methods supported, and the type of mobility of the user, etc., were not taken into consideration. Scenarios where the AN can support multiple levels of QoS or when the user is already using a network for an earlier session were not addressed. Moreover, issues of the sensitivity of selection processes to the input data were not explored. The solution presented in this chapter is based on the architectural framework described in Chapter 2 [1] that supports current business models (e.g., roaming agreements) used in the industry. It identifies key parameters for use in the network selection process, proposes a comprehensive network selection mechanism using these parameters as inputs, and then validates the proposed mechanism by applying it under different scenarios.

3.3 Decision Process in Network Selection

In the network-assisted mechanism proposed in [1] and described in Chapter 2, the network assists the terminal in the selection process by performing data collection and analysis to provide the network ranking. The architecture proposes the use of three network-based functional entities: 1) Data Collection Node (DCN) for retrieving network characteristics, 2) Service Announcement Node for providing services related data and 3) enhanced Authentication, Authorization and
Accounting (AAA) Node for AAA information. Together they provide input data to a network-based assessment entity that uses an algorithm to calculate the network rankings for use by the terminal in network selection. As indicated in the policy-based network selection example of Chapter 2, the execution of the network selection algorithm is assumed to reside in the enhanced PDF entity. This assumption holds for the rest of thesis.

The process starts with the terminal trying to get connectivity with whichever AN it can access (with or without authentication). This initial network access is similar to getting directory assistance over today's circuit-switched voice network. The initial AN may not be the optimal network for the requested service and can be changed once sufficient information has been collected to make a decision. The terminal provides its location and any other information that could be considered by the network in the analysis, e.g., service/QoS being requested, scans of the locally available PLMN IDs, SSIDs, etc. Using the information collected from the DCN, SAN and AAA nodes and those provided by the terminal, the enhanced PDF calculates the rankings of the available ANs and provides them to the terminal. The remainder of this chapter describes a comprehensive algorithmic approach to be used by the network-based assessment entity in the proposed architectural framework.

We formulate the network selection decision process in a heterogeneous networking environment as a MADM problem that deals with the evaluation of a set of alternatives using a set of attributes. The alternatives represent different choices. The decision process ranks these alternatives in order of preference using the set of attributes that provide different aspects by which the alternatives can be viewed.

The decision about network selection is dependent on several factors. In general the factors or attributes impacting network selection can be divided into following categories.

- Category 1 attributes include parameters that are not QoS-related. These parameters do not change very often and therefore can usually be provisioned in the network.

- Category 2 includes mostly the QoS-related attributes, both dynamic ones that need to be frequently updated as well as largely static ones that can be provisioned.

Table 3-1 lists the attributes that will be taken into consideration for network selection using the algorithms proposed in this chapter and considered in the subsequent chapters and classifies them according to the two categories described above.
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Abbrev</th>
<th>Brief Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operator Name</strong></td>
<td>ON</td>
<td>This attribute indicates the identity of the operator (e.g., roaming partner SSID or PLMN ID) for which the rest of the information is being provided.</td>
</tr>
<tr>
<td><strong>Authentication Mechanism</strong></td>
<td>AM</td>
<td>This attribute indicates the authentication mechanism used by the roaming partner. Examples could be SIM or user ID/password.</td>
</tr>
<tr>
<td><strong>Access Technology</strong></td>
<td>AT</td>
<td>This attribute indicates the wireless access technology that the AN uses, e.g., UMTS, WiMAX, WLAN. More specifics about the technology such as frequency of operation may also be included.</td>
</tr>
<tr>
<td><strong>Services Available</strong></td>
<td>SA</td>
<td>This attribute provides a list of supported services, e.g., 3GPP services such as IMS based VoIP.</td>
</tr>
<tr>
<td><strong>Geographic Location</strong></td>
<td>GL</td>
<td>This attribute provides information on the geographic location of the base station or AP.</td>
</tr>
<tr>
<td><strong>Coverage Area</strong></td>
<td>CA</td>
<td>This attribute provides a measure of the coverage area, for example, hotspot physical address. Due to propagation conditions, wireless coverage area can be irregular. Other attributes such as transmit signal power could be more useful in many instances.</td>
</tr>
<tr>
<td><strong>Cost per Byte</strong></td>
<td>CB</td>
<td>This attribute indicates relative transport cost of the operator for a particular access network. It would take into account factors such as use of licensed / unlicensed spectrum, roaming agreements, etc.</td>
</tr>
<tr>
<td><strong>Total Bandwidth</strong></td>
<td>TB</td>
<td>This attribute indicates how much bandwidth is available overall on the wireless access link.</td>
</tr>
<tr>
<td><strong>Allowed Bandwidth</strong></td>
<td>AB</td>
<td>This attribute indicates the bandwidth allowed by the AN on a per user basis.</td>
</tr>
<tr>
<td><strong>Utilization</strong></td>
<td>U</td>
<td>This attribute provides a measure of the current utilization of the wireless link.</td>
</tr>
<tr>
<td><strong>Packet delay</strong></td>
<td>D</td>
<td>This attribute gives the average packet delay within the access system. This is not the end-to-end delay.</td>
</tr>
<tr>
<td><strong>Packet Jitter</strong></td>
<td>J</td>
<td>This attribute measures the average delay variations within the access system. A large jitter could result in packet reordering or dropping of real-time packets at the receiver.</td>
</tr>
<tr>
<td><strong>Packet Loss</strong></td>
<td>L</td>
<td>This attribute measures the average packet loss rate within the access system.</td>
</tr>
</tbody>
</table>
3.4 Multi Attribute Decision Making (MADM)

The proposed decision process for network selection can be described by the steps shown in Figure 3.1. The process shown employs both non-compensatory and compensatory MADM algorithms, which are described in more detail below.

3.4.1 Non-Compensatory MADM Algorithms

In this class of MADM algorithms, advantages in one type of attribute cannot be traded for disadvantages in another type of attribute. This class of methods, as described in [11], include dominance, conjunctive, disjunctive and sequential elimination (e.g., lexicographic, elimination by aspects). Of particular interest for network selection are conjunctive and disjunctive methods. These two methods are not used to select a particular alternative but to separate the given alternatives into acceptable and unacceptable groups. All alternatives are considered acceptable so long as they satisfy the minimum cutoff criteria. In the case of the conjunctive method, the acceptable alternative has to meet the minimum cutoff for all the attributes under consideration. For the disjunctive method, an alternative is acceptable if one or more of the attributes meet or exceed the cutoff value for the attributes under consideration.
The two methods can be used together in decision making processes. In the case of network selection, the attributes listed in category 1 of Table 3-1 will be used as input to the conjunctive and disjunctive decision processes in order to categorize the available networks into acceptable and unacceptable alternatives for service delivery. Using the attributes from Category 1, an acceptable AN that could be further explored for suitability should satisfy all of the following:

- it supports at least one of the authentication mechanisms supported by the user/terminal, and
- it supports the access technology supported by the user/terminal, and
- it provides access to the service the user is requesting, and
- it is available in the geographic location the user is present, and
- it provides the coverage level that might have been indicated by the user's mobility profile.

The "and" relationship in the decision process between each of the attributes is an application of conjunctive MADM. The disjunctive MADM algorithm is applied to account for the fact that for many of these attributes, there are more than one alternative values and for the network to be acceptable it needs to support only one of these values.

### 3.4.1.1 Decreasing Inter-Technology Handoffs

Since the network capabilities as well as the associated transport costs would vary across different access technologies, it could be a preferred policy to initially select the network with an aim to decrease or avoid if possible potential inter-technology handoffs during an active session. Use of a non-compensatory MADM algorithm in the decision process can help decrease inter-technology handoffs. The user mobility profile can indicate the expected type of user mobility during a session. It is possible to have the user/terminal provide its mobility profile in step 1 of Figure 3.1 that will be used in the non-compensatory MADM part of the decision process along with coverage area information about the network. The user can set his profile to be "mobile" or "nomadic" and accordingly the non-compensatory part of the MADM based network selection algorithm can decide to pick only wide-area cellular networks or wireless local area networks. This is illustrated in Figure 3.2 where the two terminals are identical in every aspect and have access to both WWAN and WLAN. However because of the difference in settings of their
mobility profiles ("nomadic" vs. "mobile"), the "nomadic" terminal/user selects the WLAN while the "mobile" terminal/user selects the WWAN.

Figure 3.2 By using non-compensatory MADM algorithm, different mobility profile settings can provide selection of a different network

3.4.2 Compensatory MADM Algorithms

Application of a compensatory MADM algorithm involves the following steps:

- Identify all the alternatives and the compensatory MADM attributes impacting the decision process,

- Assign relative importance in the decision making process to each of the attributes, and

- Use a compensatory MADM algorithm to get a ranking of the alternatives.

Now we consider a general case where the selection has been narrowed down using the non-compensatory MADM algorithm to \( N \) access networks as shown in Figure 3.1. In order to apply a compensatory MADM algorithm to facilitate network selection, we consider the Category 2 attributes in Table 3-1. Candidate network \( i, NW_i \), from a decision making perspective can be represented by the following vector of network attributes,
We can represent the $N$ networks to be considered in the selection process by a matrix of their network attribute vectors as follows,

\[
NW_i = [CB_i \ TB_i \ AB_i \ U_i \ D_i \ J_i \ L_i]
\]

The following section describes the use of TOPSIS, a compensatory MADM algorithm to objectively decide on the best suited solution.

3.4.2.1 Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

The TOPSIS algorithm is based on the assumption that the chosen solution has the shortest distance from the best solution but the longest distance from the worst solution. The following steps are involved in the application of TOPSIS to the network selection problem.

1. The value for each of the attribute in the matrix is normalized. For example, entries in the second column of the matrix can be normalized as follows,

\[
(TB_{norm})_i = \frac{TB_i}{\sqrt{\sum_{f=1}^{N} TB_f^2}}
\]

2. The matrix is updated with these normalized values as follows.

\[
(NW_{norm})_i = \begin{bmatrix}
(CB_{norm})_1 & (TB_{norm})_1 & (AB_{norm})_1 & (U_{norm})_1 & (D_{norm})_1 & (J_{norm})_1 & (L_{norm})_1 \\
(CB_{norm})_2 & (TB_{norm})_2 & (AB_{norm})_2 & (U_{norm})_2 & (D_{norm})_2 & (J_{norm})_2 & (L_{norm})_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
(CB_{norm})_N & (TB_{norm})_N & (AB_{norm})_N & (U_{norm})_N & (D_{norm})_N & (J_{norm})_N & (L_{norm})_N
\end{bmatrix}
\]

3. The next step is to decide on the relative importance of each of the attributes involved in the decision about network selection. For this purpose, each of the attribute is assigned a weight “$w$”, such that
\[ W = w_{CB} + w_{TB} + w_{AB} + w_{U} + w_{D} + w_{J} + w_{L} = 1 \]

4. Using these assigned weights, the matrix from step 1 is updated as follows.

\[
\begin{bmatrix}
w_{CB} \cdot (CB_{\text{norm}}) & w_{TB} \cdot (TB_{\text{norm}}) & w_{AB} \cdot (AB_{\text{norm}}) & w_{U} \cdot (U_{\text{norm}}) & w_{D} \cdot (D_{\text{norm}}) & w_{J} \cdot (J_{\text{norm}}) & w_{L} \cdot (L_{\text{norm}}) \\
w_{CB} \cdot (CB_{\text{norm}}) & w_{TB} \cdot (TB_{\text{norm}}) & w_{AB} \cdot (AB_{\text{norm}}) & w_{U} \cdot (U_{\text{norm}}) & w_{D} \cdot (D_{\text{norm}}) & w_{J} \cdot (J_{\text{norm}}) & w_{L} \cdot (L_{\text{norm}}) \\
w_{CB} \cdot (CB_{\text{norm}}) & w_{TB} \cdot (TB_{\text{norm}}) & w_{AB} \cdot (AB_{\text{norm}}) & w_{U} \cdot (U_{\text{norm}}) & w_{D} \cdot (D_{\text{norm}}) & w_{J} \cdot (J_{\text{norm}}) & w_{L} \cdot (L_{\text{norm}}) \\
\end{bmatrix}
\]

5. The next step is to find the best and worst value for each of the attribute. Depending on the attribute, the best (or the worst) value can be either the maximum or the minimum value. For example, in the case of the network utilization attribute, the best value will be the lowest one and the worst value will be the highest one. For the case of the allowed-bandwidth attribute, however, the best value will be the highest one and the worst value will be the lowest one. In the case of TB, this calculation in general can be represented mathematically as follows. Based on the description above, these equations would evaluate to best value for TB being the maximum and the worst value for TB being the minimum.

\[
TB_{\text{Best}} = \max(TB_{\text{w-norm}}) \quad |i = 1,2,\ldots,N \\
TB_{\text{Worst}} = \min(TB_{\text{w-norm}}) \quad |i = 1,2,\ldots,N
\]

6. For each of the access network under consideration (represented by a row in the matrix), the measure of separation, both for the best and worst cases, is calculated

\[
S_{\text{Best}} = \frac{((CB_{\text{w-norm}})_i - CB_{\text{Best}})^2 + ((TB_{\text{w-norm}})_i - TB_{\text{Best}})^2 + ((AB_{\text{w-norm}})_i - AB_{\text{Best}})^2 + ((U_{\text{w-norm}})_i - U_{\text{Best}})^2 + ((D_{\text{w-norm}})_i - D_{\text{Best}})^2 + ((J_{\text{w-norm}})_i - J_{\text{Best}})^2 + ((L_{\text{w-norm}})_i - L_{\text{Best}})^2}{\sqrt{((CB_{\text{w-norm}})_i - CB_{\text{Best}})^2 + ((TB_{\text{w-norm}})_i - TB_{\text{Best}})^2 + ((AB_{\text{w-norm}})_i - AB_{\text{Best}})^2 + ((U_{\text{w-norm}})_i - U_{\text{Best}})^2 + ((D_{\text{w-norm}})_i - D_{\text{Best}})^2 + ((J_{\text{w-norm}})_i - J_{\text{Best}})^2 + ((L_{\text{w-norm}})_i - L_{\text{Best}})^2}}
\]

\[
S_{\text{Worst}} = \frac{((CB_{\text{w-norm}})_i - CB_{\text{Worst}})^2 + ((TB_{\text{w-norm}})_i - TB_{\text{Worst}})^2 + ((AB_{\text{w-norm}})_i - AB_{\text{Worst}})^2 + ((U_{\text{w-norm}})_i - U_{\text{Worst}})^2 + ((D_{\text{w-norm}})_i - D_{\text{Worst}})^2 + ((J_{\text{w-norm}})_i - J_{\text{Worst}})^2 + ((L_{\text{w-norm}})_i - L_{\text{Worst}})^2}{\sqrt{((CB_{\text{w-norm}})_i - CB_{\text{Worst}})^2 + ((TB_{\text{w-norm}})_i - TB_{\text{Worst}})^2 + ((AB_{\text{w-norm}})_i - AB_{\text{Worst}})^2 + ((U_{\text{w-norm}})_i - U_{\text{Worst}})^2 + ((D_{\text{w-norm}})_i - D_{\text{Worst}})^2 + ((J_{\text{w-norm}})_i - J_{\text{Worst}})^2 + ((L_{\text{w-norm}})_i - L_{\text{Worst}})^2}}
\]

7. For each of the access networks under consideration (represented by a row in the matrix), its level of preference is measured. The preference level “P” of network i measured in terms of distances “S” from the best and worst solutions, respectively $S_{\text{Best}}$ and $S_{\text{Worst}}$, is given by

\[
P_i = \frac{S_{\text{Worst}}}{S_{\text{Best}} + S_{\text{Worst}}}
\]
The preference value "P" represents an indifference curve. The term indifference curve means that a decision maker would give equal preference to any of the alternatives on the same indifference curve, i.e., with the same value of "P". Figure 3.3 illustrates some indifference curves for a system with two attributes: cost per byte and throughput.

8. The access network with the highest "P" value is selected.

![Indifference Curves](image)

**Figure 3.3 Example of indifference curve for TOPSIS while using attributes of cost per byte and throughput**

The computational complexity involved in calculating the Euclidean distances used in TOPSIS is minimal. Therefore the computational time and the resources needed for it should not be significant. Besides, in the architectural framework defined in Chapter 2 [1], data collection as well as the analysis involving MADM are performed in the network and only the network rankings are conveyed to the terminal over the wireless link. This approach is resource efficient from the wireless bandwidth utilization perspective.

3.4.2.2 Assignment of Attribute Weights

The assignment of weights to different parameters in a compensatory MADM algorithm plays a key role in network selection. It is proposed that the assignment of these weights be based on the
interpretation of the requested service/application requirements or the subscribed user profile by
the subscriber’s HN operator. Two possible types of QoS profiles that can be stored in the user’s
HN are thus as follows:

- An overall user QoS profile that is applicable to all of the services that the user is using;
e.g., Gold, Silver, or Bronze profile can indicate the level of QoS that the user is expected
to have based on the subscription.

- A QoS profile of an individual service that is applicable to all subscribers of that service;
e.g., VoIP service profile or web browsing profile.

The relative importance of different attributes for some common types of services/applications or
user subscriptions is described below.

**VoIP** – This is a low bandwidth application that is very sensitive to delay and jitter but can
withstand some packet losses. The transport cost factor is considered negligible because of low
bandwidth usage. Also because of low bandwidth requirements, the total bandwidth and available
bandwidth are not significant factors. Since there is some correlation between the utilization and
jitter / delay, it is preferred to have a low utilization for the selected network.

**Streaming** – Being a multimedia service, a streaming application requires a higher bandwidth than
VoIP. Therefore the available bandwidth, transport cost and current utilization are important
factors. The service is less vulnerable to delay and jitter than VoIP because of its ability to buffer
longer duration of data before play back. The sensitivity to packet loss is similar to VoIP, such
that some packet losses can be compensated without impact to user experience.

**Web Browsing** – Web browsing type applications require a low QoS service; i.e., the importance
of utilization, delay, jitter, and packet loss is low. It does not need a guaranteed bit rate because of
the bursty nature of web traffic pattern. With statistical traffic multiplexing for such type of
traffic, broadband wireless networks can deliver a reasonable customer experience even at lower
average data rates. The total bandwidth and allowed bandwidth are therefore less critical but the
transport cost is considered critical.

**Gold Subscription** – This indicates a premier user subscription that would allow the use of the
highest level QoS independent of the transport cost.

**Silver Subscription** – This indicates a medium priority user subscription that would try to balance
between the QoS requirements and other factors such as the transport cost.
Bronze Subscription – This indicates a lower priority user subscription where the transport cost is significantly important compared with any QoS parameters.

3.4.3 Results of Use Case Scenarios

To validate the use of TOPSIS in the decision process we ran the decision process for network selection among five candidate networks. Table 3-2 provides the numerical attribute values for these five networks that were used for illustrative purposes in the decision process for network selection. They are representative of the listed example network types that a typical user could expect. For example, the cost attribute is derived on the basis of spectral efficiency of the technology and whether the technology runs on licensed or unlicensed spectrum. So the cost is lowest for an unlicensed spectrally efficient technology such as IEEE 802.11n and it is highest for a licensed and relatively less spectrally efficient technology such as UMTS. Also the maximum estimated throughput for each of the example technologies has been used for the total bandwidth attribute. The allowed bandwidth has been assumed to be based on operator policy to rate limit its customers differently for different access technologies. Other QoS related attributes such as delay, jitter and loss represent a snapshot of the values that could exist in these networks at the time of decision.

Table 3-2 Attribute values for Scenarios 1, 2, 3 & 4

<table>
<thead>
<tr>
<th>Ntwk#1 e.g. UMTS</th>
<th>CB (%)</th>
<th>TB (mbps)</th>
<th>AB (mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10⁶)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>2</td>
<td>0.2</td>
<td>10</td>
<td>400</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Ntwk#2 e.g. 802.11b</td>
<td>20</td>
<td>11</td>
<td>1</td>
<td>20</td>
<td>200</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Ntwk#3 e.g. 802.11a</td>
<td>10</td>
<td>54</td>
<td>2</td>
<td>20</td>
<td>100</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Ntwk#4 e.g. 802.11n</td>
<td>5</td>
<td>100</td>
<td>5</td>
<td>40</td>
<td>150</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Ntwk#5 e.g. 4G</td>
<td>30¹, 5²</td>
<td>100</td>
<td>5</td>
<td>20</td>
<td>100</td>
<td>20</td>
<td>15</td>
</tr>
</tbody>
</table>

1- Scenarios 1, 2.
2- Scenarios 3, 4

We considered the following use case scenarios where the service and user subscription profiles as described previously were used.
**Scenario 1** – In this scenario the decision process is influenced by the requested service indicated by the user. The service type VoIP, streaming and web browsing were used to assign attribute weights.

**Scenario 2** – In this scenario the decision process is influenced by the QoS profile of the user stored in the HN. The user subscription type of Gold, Silver and Bronze were used to assign attribute weights.

**Scenario 3** – This is the same as scenario 1 except that the operator of Ntwk #5 has temporarily indicated a reduction of access cost to the subscriber’s HN in order to attract more customers to its network. This change would influence the rankings.

**Scenario 4** – This is the same as scenario 2 except that the operator of Ntwk #5 has temporarily indicated a reduction of access cost to the subscriber’s HN in order to attract more customers to its network.

---

**Figure 3.4 Results for scenarios 1 & 3**

57
The results for scenarios 1 and 3 are shown in Figure 3.4, which also shows the weights assigned to various attributes. The assignments are based on the analysis of the service/subscribed QoS profiles described earlier. Similarly the results for scenarios 2 and 4 are documented in Figure 3.5. The results indicate that even with the same set of parameter values, the top candidate network can be different for different types of QoS subscriptions or service profiles; e.g., the network considered optimal for a VoIP user is different than that for a web browsing user in scenario 1. This is due to differences in the service requirements reflected by weights assigned to the parameters for QoS profiles for different services.

The network operators can influence the decision process by making their network more attractive to a certain type of users or services. In the case of scenarios 3 and 4, e.g., Ntwk #5 influences the selection by temporarily decreasing the access cost for its network to its roaming partners. As a result when the user’s HN calculates the network rankings, Ntwk #5 becomes quite attractive for most of the services or users where the network transport cost is an important factor in the

Figure 3.5 Results for scenarios 2 & 4

The network operators can influence the decision process by making their network more attractive to a certain type of users or services. In the case of scenarios 3 and 4, e.g., Ntwk #5 influences the selection by temporarily decreasing the access cost for its network to its roaming partners. As a result when the user’s HN calculates the network rankings, Ntwk #5 becomes quite attractive for most of the services or users where the network transport cost is an important factor in the
decision process. This is evident by observing the results of Figures 3.4 and 3.5, where Ntwk #5 becomes the network of choice for Gold, Silver and Bronze customers along with Streaming Media and Web Browsing services. Also it can be observed that the network rankings for services/subscriptions that are not significantly influenced by the cost attribute (e.g., VoIP), are not impacted by changes in Ntwk #5’s cost.

3.4.4 Sensitivity of Network Selection to Dynamic Attribute Values

The compensatory part of the network selection process is primarily related to finding the network that is best suited in terms of QoS to the user or the QoS profile of the requested service. Many of the QoS related parameters used in the proposed compensatory MADM algorithm are dynamic. An inaccurate measure of these parameters values could result in the selection of a non optimal network.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>P - Ntwk#1</th>
<th>P - Ntwk#2</th>
<th>P - Ntwk#3</th>
<th>P - Ntwk#4</th>
<th>P - Ntwk#5</th>
</tr>
</thead>
<tbody>
<tr>
<td>VoIP Baseline</td>
<td>0.2912</td>
<td>0.7190</td>
<td>0.8830</td>
<td>0.6161</td>
<td>0.8570</td>
</tr>
<tr>
<td>VoIP 10% Error</td>
<td>0.2919</td>
<td>0.7272</td>
<td>0.8582</td>
<td>0.6237</td>
<td>0.8657</td>
</tr>
<tr>
<td>Web Baseline</td>
<td>0.1070</td>
<td>0.7725</td>
<td>0.8623</td>
<td>0.8901</td>
<td>0.7504</td>
</tr>
<tr>
<td>Web 30% Error</td>
<td>0.1251</td>
<td>0.7738</td>
<td>0.8638</td>
<td>0.8687</td>
<td>0.7517</td>
</tr>
</tbody>
</table>

Table 3-3 provides a comparison of the results for TOPSIS in scenario 1 as previously described, when the attribute values for delay, jitter and loss are changed by 10% for a VoIP user and when the attribute values for delay, jitter and loss are changed by 30% for a web user. The results indicate that while changing these attribute values by 10% changes the selected network for a VoIP user, even changing these attribute values by 30% does not change the result for a web user. This shows that network selection for services like web browsing or for bronze users are less
sensitive to the dynamic attribute information compared with services such as VoIP or Gold users. This is due to the assignment of lower weights to the dynamic attributes in the QoS profiles for a bronze user or web browsing. Consequently errors in these attribute values have a lesser impact on network selection and the results are more reliable.

### 3.4.5 Network Selection with Managed IP Networks

Managed IP networks typically provide multiple levels of SLAs, which can possibly be mapped to several classes of QoS. For example, the QoS classes identified by different DiffServ Codepoint markings at the IP layer can be mapped to different SLA levels. For MADM algorithms used in network selection, networks that support multiple levels of SLAs or classes of services can be mapped to multiple network options in the MADM algorithm, each supporting one QoS class or SLA.

#### Table 3-4 Attribute values for Scenarios 1, 2, 3 & 4

<table>
<thead>
<tr>
<th>QoS Class/SLA#1</th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>0.1</td>
<td>30</td>
<td>400</td>
<td>100</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>QoS Class#2/SLA#2</td>
<td>20</td>
<td>100</td>
<td>0.5</td>
<td>20</td>
<td>200</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>QoS Class#3/SLA#3</td>
<td>30</td>
<td>100</td>
<td>1.0</td>
<td>20</td>
<td>100</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>QoS Class#4/SLA#4</td>
<td>50</td>
<td>100</td>
<td>2.0</td>
<td>40</td>
<td>50</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>QoS Class#5/SLA#5</td>
<td>50</td>
<td>100</td>
<td>5.0</td>
<td>20</td>
<td>10</td>
<td>5</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3-4 represents an access system that supports five classes of service or five levels of SLAs. The attribute values for use in the decision process for network selection are given for illustrative purposes. For example, the cost attribute in this case is derived based on how the network would treat the packets from that class of service. QoS Class 1, for example, gets the least preferential treatment relative to the other classes and therefore has the lowest cost. The selection of any particular alternative will map to the same physical network but with a different QoS class. Therefore, while the total bandwidth of the AN will be the same for all alternatives, the values of other parameters (e.g., delay, jitter, packet loss) are different depending upon the QoS class. The allowed bandwidth has been assumed to be based on operator policy to rate limit its customers using different QoS classes on the access network. For TOPSIS, this access system is represented as five alternatives.
The results for the four scenarios discussed previously using TOPSIS are again exhibited in Figure 3.4 and Figure 3.5, which demonstrate how different user subscriptions or service requirements are mapped to different QoS classes or SLA levels represented by different alternatives in the TOPSIS algorithm. For example, use of TOPSIS algorithm maps a Streaming media user in Scenario 1 to QoS class/SLA level #5 while a web browsing user maps to QoS class/SLA level #1 for the same attribute values. However a decrease in the cost attribute for QoS class/SLA level #5 in scenarios 3 and 4 results in it being selected for all the service and subscription types.

3.4.6 Network Selection for Accessing Multiple Services Simultaneously

So far network selection has been discussed for scenarios where the user accesses one or more services having the same QoS requirements. A common scenario that could arise is that a user who is already in session for a service using a network previously selected for this service now decides to start another service that may have different service requirements in terms of QoS. This would trigger a new network selection decision. Although it is possible to select two different networks and use them to access two services simultaneously, in practice such a policy can create several problems such as authentication with two networks simultaneously, excessive power consumption or possible interference of turning on two radios at the same time, and appropriate routing of service data within the device and the networks. For these reasons it is recommended that such network selection solutions be avoided. The proposed solution using the comprehensive MADM algorithm based network selection is as follows:

1. Perform the non-compensatory MADM part of network selection as usual for each of the individual service being requested.

2. Use a compensatory MADM algorithm for each of the individual service requested. Networks that support multiple levels of QoS/SLAs would be mapped to more than one network options as described in the earlier section.

3. The final network selected will be based on the average of the rankings for the networks selected for the different services; e.g., a network ranked as #1 and #3 for the two services under consideration will have an average ranking of #2 for the combined service delivery.
There may also be other aspects of the already in-use service that can influence the subsequent network selection upon starting another service. For example the already in-use service can be VoIP that requires support for real-time traffic by the AN. As this may not be possible across all candidate networks, it may be a policy in such a situation to use the already in-use network for the new service as well. In addition, certain requirements of the new service (e.g., QoS) may force a network reselection for the service already in use. Such requirements would therefore influence the final network selection decision. There may still be other services which may not be possible to combine over the same wireless AN. Network selection in such a situation could be dependent upon provisioned operator or user policies that specify which service would take priority and therefore the delivery of the lower priority service would have to be terminated.

3.5 Conclusions

Selection of an optimal access network is an important aspect of service delivery in a heterogeneous wireless system. The decision is influenced by several factors. In this chapter we have proposed a decision process for a network-assisted network selection mechanism that combines non-compensatory and compensatory MADM algorithms. The proposed mechanism is more comprehensive than the work published earlier. Along with disjunctive and conjunctive non-compensatory MADMs, we have described the use of TOPSIS, a compensatory MADM algorithm. The TOPSIS algorithm has been validated by applying it to a large variety of use cases. Scenarios with different services and subscribed QoS profiles have been described to observe how the proposed mechanism would work. Application of the proposed mechanism to networks with multi-level SLAs has been described. Possible approaches to network selection when a service is already in use have been discussed and solutions are proposed. The results of this chapter provide a basis for further research into the area of service delivery in a heterogeneous networking environment.
3.6 References


4 Enhancements to MADM Algorithms for Use in Automated Network Selection where System Parameters are Known

Network convergence across different access technologies holds a promise of enabling ubiquitous service availability. However it also implies that services have to be delivered over a heterogeneous mix of access technologies. There are several technical challenges that have to be overcome in order to provide a good service experience to users in such an environment. The selection of an optimal service delivery network in the presence of multiple ANs is one of the important issues. This problem of identifying the best suited service delivery network when the user has multiple networks to choose from is called network selection. The problem is more significant for all-IP networks where service experience can seriously degrade depending upon the network capabilities and congestion levels. It is an area of active research and a topic of discussion in several standardization forums. The IEEE 802.11u Working Group currently has a draft proposal that would enable information exchange for network selection between the network and the terminal [1]. It leverages the protocol being developed by IEEE 802.21 for this purpose which they also plan to use for selecting networks for vertical handoffs [2]. Similarly 3GPP is looking into the mechanism of network discovery and selection in their work on SAE [3]. Along

---

1 Parts of this chapter have been published in the following
with the protocol and architectural aspects of the problem, an essential component in solving the problem of network selection is defining the optimization objective and the algorithm to be used in the selection process. Selection of a non-optimal network creates undesirable results such as poor customer experience or the use of a more expensive network. The focus of this chapter is on this aspect of the problem.

The previous chapter described how the problem of network selection can be solved using a comprehensive MADM approach [4]. The algorithm used for this purpose was the standard version of TOPSIS [5]. In this chapter we consider three MADM algorithms representing three different decision making philosophies and propose several improvements to these standard MADM algorithms [5] to make them more suitable to solving the problem of network selection. The rest of the chapter is divided into three main parts. The first part proposes an improvement to TOPSIS that helps remove ranking abnormalities from the results through an iterative process. In the second part we explore the application of MADM algorithms to network selection where the optimization objectives of the decision maker require the use of non-monotonic utility for some of the attributes. We show why GRA is better suited when compared with other standard forms of MADM algorithms for such scenarios. Finally in the last part of the chapter we propose a new adaptation of ELECTRE [6] to solve the problem of network selection with non-monotonic utility for some of the attributes.

The key difference between different MADM algorithms is that they are based on different decision making philosophies. This allows the decision maker to choose the algorithm that is closest to its philosophy on decision making. We have avoided any discussion on a direct comparison amongst the various algorithms discussed in the chapter on the basis of their accuracy of results. This issue has been studied in the past and the conclusion established that such comparisons are not possible as they would need the use of another MADM algorithm. However, we have compared the MADM algorithms in terms of their suitability to solving the problem of network selection.

4.1 Use of Iterative TOPSIS for Network Selection in Heterogeneous Wireless Access

In this section we analyze TOPSIS, a compensatory MADM algorithm. We look into ranking abnormalities for TOPSIS and propose an iterative approach to improve the algorithm’s application to the problem of network selection. We show that for the problem of network
selection, where ranking of only the top candidates is important, an iterative approach that removes the lowest-ranked candidate after each iteration can provide improved results. The approach proposed here can be used in the network-assisted network selection architecture described in Chapter 2 [7].

4.1.1 TOPSIS

MADM algorithms are used for determining the ranking of alternatives in terms of their desirability with respect to multiple criteria that can influence the decision. For the problem of network selection, we are only interested in identifying the top ranking candidates and this can be considered as a special application of finding the ranking of all the alternatives. In order to formulate a problem as MADM, we consider \( A = \{ A_i, \text{ for } i=1,2,\ldots,n \} \), a set of a finite number of alternatives. Also, we consider \( C = \{ C_j, \text{ for } j=1,2,\ldots,m \} \) to be a set of attributes against which the alternatives are to be judged, and \( w_1, w_2, \ldots, w_m \) are the weights that represent the relative importance of these attributes. Then the MADM problem can be represented by a matrix shown in Figure 4.1.

TOPSIS is a widely used MADM algorithm developed by Hwang and Yoon [8]. It is applicable to problem spaces in which the attributes have monotonically increasing or decreasing levels of utility. The algorithm calculates perceived positive and negative ideal solutions based on the range of attribute values available for the alternatives. The premise of the algorithm is that the best solution is the one with the shortest distance to the positive ideal solution and longest distance from the negative ideal solution, where distances are measured in Euclidean terms. All
solutions with the same preference level can be mapped to an “indifference curve”, which shows how different values of the attributes when combined can result in the same level of preference.

For the network selection problem, the representative set of attributes considered in the decision making process are Cost per Byte (CB), Total Bandwidth (TB), Allowed Bandwidth (AB), Utilization (U), Packet Delay (D), Packet Jitter (J) and Packet Loss (L). Their descriptions are provided in Table 3-1 of Chapter 3.

Using the attributes defined above, a candidate network NW for selection using MADM can be represented by the attribute vector:

\[ NW = [CB \ TB \ AB \ U \ D \ J \ L] \]

If there are \( N \) network alternatives to be considered in the selection process, their attributes can be represented in the form of a matrix as follows,

\[
NW = \begin{bmatrix}
CB_1 & TB_1 & AB_1 & U_1 & D_1 & J_1 & L_1 \\
CB_2 & TB_2 & AB_2 & U_2 & D_2 & J_2 & L_2 \\
& & & & & & \\
& & & & & & \\
& & & & & & \\
& & & & & & \\
\end{bmatrix}
\]

Now we apply the TOPSIS algorithm to the problem in a stepwise manner as follows. The detailed steps have been described in the previous chapter. Here we summarize them as they will be referred in the chapter during analysis of ranking abnormalities.

1. The value for each of the attribute in the matrix is normalized using Euclidean normalization.
2. The matrix is updated with these normalized values.
3. The relative importance of each of the attributes involved in the decision about network selection is determined and represented by a weight.
4. Using these assigned weights, the matrix from step 1 is updated.
5. The best and worst values for each of the attribute are determined. Depending on the attribute, the best (or worst) value can be either the maximum or the minimum value. For example, in the case of the network utilization attribute, the best value will be the lowest and worst value will be the highest. For the case of an attribute related to the allowed bandwidth, however, the best value will be the highest and worst value will be the lowest.
6. For each of the access network under consideration, the measures of closeness/separation “S”, for the best and worst cases, are calculated using Euclidean distances.

7. For each of the access networks under consideration (represented by a row in the matrix), its level of preference is measured. The preference level “P”, measured in terms of distances “S” from the best and worst solutions, is represented by the following formulation

\[ P_i = \frac{S_{\text{Worst}}}{S_{\text{Best}} + S_{\text{Worst}}} \]

8. The access network with the highest “P” value is selected.

4.1.2 Ranking Abnormalities

MADM algorithms are known to suffer from ranking abnormalities and TOPSIS is not any different. Triantaphyllou [9] has done significant work in documenting this problem for MADM algorithms. He defines an effective MADM algorithm to be one for which the indication of the best alternative does not change when an alternative that is not the best is replaced by another worse alternative, given that the relative importance of each decision criterion remains unchanged [9]. However, MADM algorithms do not always exhibit this ideal behavior.

It has been shown that the frequency of such ranking abnormalities depends upon the algorithm as well as the problem space (e.g., the general spread of data values for different attributes). In general, these abnormalities increase with the number of attributes and also with the number of alternatives under consideration [10][11].

In order to better understand the occurrence of ranking abnormalities in TOPSIS, we make use of a graphical representation of the method. Figure 4.2 graphically represents a decision making scenario using TOPSIS with 4 alternatives, and two attributes. The alternatives are mapped into a two dimensional space that represents the two attributes. The attribute data used have been normalized and adjusted based on attributes’ importance as described by steps 1 through 4 earlier. The figure also shows the Euclidean distances of alternatives #4 and #2 from the positive and negative ideal solutions, which are calculated as described by step 5, while the distances are calculated per step 6. The two alternatives are equidistant from the positive and negative ideal solutions and therefore equal in their rankings. It is also clear from the figure that alternative #3 has the lowest ranking because of its relative closeness to the negative ideal solution as compared
to the positive ideal solution. Similarly alternative #1 is at the top of the ranking because of its relative closeness to the positive ideal solution when compared to the negative ideal solution.

Now, if the lowest ranking alternative is removed, for an ideal MADM algorithm, the comparison of the remaining alternatives should not be impacted. However this is not always the case and ranking based on TOPSIS can be impacted [9]. For reasons explained later in the section, with the removal of alternative #3 from Figure 4.2, both positive and negative ideal solutions as well as the normalized attribute values for the remaining alternatives will be changed. In the case of network selection, we are only interested in the top ranking candidate networks. For an ideal MADM algorithm, inclusion of an alternative of lowest ranking in the comparison should not have any impact on the ranking of top ranking alternatives. If the inclusion of a lowest ranking alternative changes the rankings amongst top ranking alternatives, then it indicates the presence of ranking abnormality. In this situation, when we are interested in only the top ranking alternative, the removal of a candidate that causes ranking abnormality can provide more reliable results.

**Figure 4.2 Use of TOPSIS for a MADM problem with 4 alternatives and 2 attributes**

4.1.2.1 Factors contributing to ranking abnormality for TOPSIS

Attributes used in TOPSIS are likely to be represented in different units of measurements (e.g., dollars for transport per byte cost or milliseconds for latency). Before attribute values can be used for calculating positive and negative ideal solutions, they have to be expressed in the same units or made dimensionless. The attributes are made dimensionless by the process of normalization. There are more than one methods of normalizing the data [8]. A Euclidean normalization
technique is used in TOPSIS as shown in step 1, and removing or adding an alternative will change the normalized attribute values for all the candidate networks. Depending upon the attribute values of the removed alternative, the impact on the normalized values for different attributes of the remaining alternatives will be different. Subsequently, the calculation of weight-adjusted normalized values of step 4 will also be impacted non-linearly. As a result, the best and worst values for each of the attributes in step 5 will also be changed non-linearly. \cite{11} describes the impact on ranking of alternatives to changes in the values of attributes by performing sensitivity analysis. It shows that ranking order can change if the values of the attributes in relationship to others are changed by a certain extent. The effect of removing an alternative from a comparison is similar to changing the attribute values. Depending upon the attribute values of the alternative being removed, the sensitivity and hence chances of a change in the rankings of the remaining alternative being considered can vary.

For ranking purposes, TOPSIS uses the Euclidean distance of attributes from their respective positive and negative ideal values (as described in step 6). After the removal/addition of an alternative from the bottom of the ranking, the Euclidean distance calculations for an alternative will be based upon the new normalized attribute values and the new best and worst values. As described previously, the relative change in these new values from the old values can be different for different alternatives. Therefore the relative change in the new Euclidean distance from its prior value for any one alternative could also be different than for other alternatives. As a result, the calculation of preference levels \("P\", as defined in step 7 can in this situation provide a different ranking order than the prior one.

\subsection*{4.1.3 Iterative MADM Approach using TOPSIS}

Because of the nature of MADM problems, it is not possible to objectively measure the accuracy of rankings obtained from MADM algorithms. For MADM algorithms that can exhibit non-ideal behaviors, there is a possibility that the inclusion of a bottom ranked alternative can alter the rankings of the top ranked alternatives. If instances of such abnormalities can be identified, then for some problem spaces it is possible to take action to remove their impact. For example, for MADM algorithms which exhibit such a non-ideal behavior and the decision maker is interested only in the top ranking alternatives, then a comparison amongst only the likely top candidates will provide more reliable rankings than a comparison involving all of the candidates including those
at the bottom of the ranking. This is especially true if the inclusion of a bottom ranked alternative causes a change in the ranking at the top.

The factors identified in the previous sections are not known to cause radical changes to ranking; e.g., it is unlikely to cause the top candidate to appear at the bottom of the ranking list. While it is not possible to remove causes of ranking abnormalities in TOPSIS, the fact that in the case of network selection our interest is only in the top candidate networks can be used to improve the accuracy of the results. It is therefore possible to adopt an iterative approach whereby, in each of the iteration, the network at the bottom of the ranking is removed and the rest of the alternatives are compared. For most use case scenarios, it is expected that the total number of candidate networks in a geographic location would be in the order of ten or less, and therefore the iterative approach will continue to be efficient. For a larger number of alternatives, it is possible to use a variation of the proposed approach where each time the bottom quartile (or half) of the alternatives are removed from the next comparison. The process as shown in Figure 4.3 is repeated till only the top candidate networks are left. If in the final iteration the top candidates are very close in their ranking index, then it is proposed to use one of the attribute considered critical by the decision maker as a tie breaker. For example, if the network operator is the decision maker, then cost can be the decision factor for deciding between two equally suitable networks and hence used as a tie breaker.

**Figure 4.3 Decision making process in Iterative TOPSIS approach**
4.1.3.1 Results for use case scenario

In order to validate the use of the iterative TOPSIS approach as shown in Figure 4.3, five networks were considered in our decision process. For illustrative purposes, Table 4-1 presents numerical attribute values for these networks at the time of network selection. They are representative of listed example network types that a typical user could expect and Section 3.4.3 provides explanation for the assignment of these attribute values.

The weights assigned to the attributes used in TOPSIS were based on a subscribed QoS profile stored in the user’s HN. A Bronze subscription was assumed in this example which indicates an overall lower priority user subscription where transport cost is significantly important compared with any QoS parameters for the candidate network to be selected. The actual weights assigned to the attributes per step 3 are shown in Figure 4.4. Figure 4.4 shows that whereas the cost per byte...
is given a very high weight, the allowed bandwidth and total bandwidth have been assigned a weight of zero, indicating that they do not play any role in network selection for a Bronze user.

Table 4-2 Results of Iterative approach for TOPSIS

<table>
<thead>
<tr>
<th>No. of iteration</th>
<th>P - Ntwk#1</th>
<th>P - Ntwk#2</th>
<th>P - Ntwk#3</th>
<th>P - Ntwk#4</th>
<th>P - Ntwk#5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1180</td>
<td>0.8371</td>
<td>0.9326</td>
<td>0.8808</td>
<td>0.7368</td>
</tr>
<tr>
<td></td>
<td>Rank - 5</td>
<td>Rank - 3</td>
<td>Rank - 1</td>
<td>Rank - 2</td>
<td>Rank - 4</td>
</tr>
<tr>
<td>2</td>
<td>0.4101</td>
<td>0.8022</td>
<td>0.8854</td>
<td>0.1158</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rank - 3</td>
<td>Rank - 2</td>
<td>Rank - 1</td>
<td>Rank - 4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.1182</td>
<td>0.6719</td>
<td>0.8759</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rank - 3</td>
<td>Rank - 2</td>
<td>Rank - 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>0.1886</td>
<td>0.8114</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rank - 2</td>
<td>Rank - 1</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-2 shows the results of the proposed iterative implementation of TOPSIS. The algorithm was run four times, each time removing the candidate network at the bottom of the ranking. On the first iteration it is observed that the ranking of the top two candidate networks is reversed indicating that network at the bottom had some influence on the overall ranking. Further iterations, however, do not change the ranking any further. The results therefore show that removal of the alternative at the bottom of the initial ranking from subsequent comparisons can result in the stabilization of the top ranked alternative. For an ideal MADM algorithm, the inclusion or exclusion of a bottom ranking alternative should not change the ranking at the top. The results of the first iteration indicate that the initial top rankings obtained using TOPSIS were not reliable as they were influenced by the lowest ranking alternative. Therefore TOPSIS did not exhibit an ideal behavior in this case. Comparing only the more likely candidates by removing the bottom ranked candidate (that would not have been selected) after the first iteration removed this ranking abnormality and provided more reliable results. It should be noted that it is not possible to objectively measure the accuracy of ranking obtained by an MADM algorithm and it is possible that two different MADM algorithms provide different rankings with the same attribute related data and decision criteria. The iterative TOPSIS approach described in this section is, however,
capable of overcoming ranking inconsistencies when the TOPSIS algorithm is used to select the top candidate.

4.2 Utility Aspects of Attributes in Application of MADM Algorithms to Network Selection

In this section we view various MADM algorithms for their application to network selection in the light of optimization objectives of a decision maker requiring the use of non-monotonic utility for some of the attributes. We show why GRA is better suited when compared with other standard forms of MADM algorithms for such scenarios.

4.2.1 Deficiency in Current Studies

The need to provide a consistent service experience requires selection of optimal delivery network. The topic is of special interest for multimode IP devices where services can be delivered over a variety of wireless access technologies under varying network conditions. Several factors related to network capabilities and QoS conditions influence the network selection decision process, e.g., bandwidth, delay, jitter, and packet loss. This makes the use of deterministic decision making tools such as MADM algorithms [8] possible. Their use has been previously considered, e.g., for network selection in a heterogeneous wireless networking environment [12][13][14][15], to derive a ranking of the available networks in terms of their suitability. The highest ranking network is then selected as the best suited network. The prior work however failed to provide a comparison amongst the MADM algorithms for use in network selection. This has primarily been because of the fact that to provide a comparison in terms of ranking accuracy one would need another MADM algorithm which creates a paradoxical situation. Nonetheless it is possible to evaluate these algorithms in terms of their appropriateness for application to the problem space of network selection. So far it has been difficult to perform such an evaluation of the algorithms because of a rather simplistic assumption about optimization objectives of the decision maker. Prior studies ignored a possibly diverse range of optimization scenarios based on service and user types that could exist and hence can help in comparing the suitability of the algorithms.

The following section describes the concept of non-monotonic utility for the attributes that can help meet a wider variety of optimization objectives. We propose that this concept be leveraged in assessing the suitability of MADM algorithms to network selection.
4.2.2 Use of Non-monotonic Utility for Attributes in Network Selection

In general, as part of the decision process, the MADM algorithms associate a measure of suitability or appropriateness, hereafter called utility, with the individual attribute’s value. The utility is said to be monotonic if the measure of suitability associated with the attribute shows a monotonic increase or decrease with an increase in the attribute value. Otherwise it is said to be non-monotonic. Figure 4.5 shows a simple decision making scenario with one attribute, namely delay, and two networks. The delay attribute is shown with possible monotonic and non-monotonic utility for different service types. The monotonic utility represents optimization objectives where the network with the least delay value, i.e., Ntwk #1 will be selected for all service types. The non-monotonic utility of the delay attribute in Figure 4.5 represents the optimization objective of the decision maker where it would like to use the network that provides a delay closest to the service’s delay requirement and not necessarily the network with the least delay. This type of optimization objective would result in the selection of Ntwk#1 for VoIP service and Ntwk#2 for streaming media and web browsing services. The decision maker may desire this type of optimization for policy reasons such as load balancing across different ANs or for keeping the best network for higher QoS services and sessions that it can expect to serve. It would be similar to the policy of an airline that decides to fly with some first or business class seats empty and not to upgrade people from the economy class, with the knowledge or the hope that it would be able to fill those seats with full fare business or first class customers at the next stop.

In the case described above only one attribute was considered. For the case of multiple attributes, the overall ranking is either obtained via adding the utility associated with each of the attributes or by comparing the utilities for the attributes individually in the decision process. In prior applications of MADM to the network selection problem [12][13][14][15], the impact of different optimization objectives and hence the use of non-monotonic utility in the decision process were not studied and a general assumption about the use of the best network irrespective of service requirements or user type were considered. This implies a monotonic utility for all attributes. While some of the decision related attributes can be considered to have monotonically increasing or decreasing utilities, in reality the overall optimization goals of the decision maker may require a combination of monotonic and non-monotonic utilities for different attributes that are taken into consideration during the decision process for network selection. Associating a monotonic

76
increasing or decreasing utility in general with all attributes is therefore a simplistic assumption that would limit the scope of the types of optimization available to the decision maker [13][15].

An example of an optimization objective can be to find the network that along with other factors (such as cost, etc.) also has the best QoS characteristics from among the list of available networks. In this case the utility of QoS attributes can be considered to be monotonic. However, under a different deployment scenario, the decision maker may wish to assume a non-monotonic utility for some of the attributes considered in the selection process. An example would be an optimization objective to distribute network traffic across different access networks by selecting the access network offering a QoS closest to that being requested by the service, and not the network that may have the best QoS that far exceeds the service's QoS requirements.

Here we analyze the suitability of some of the commonly used MADM algorithms for the problem of optimal network selection, where not all the attributes considered in the decision making process have a monotonically increasing or decreasing utility. Such network selection scenarios will be quite common in future heterogeneous wireless networking environments used for delivery of both real time and non-real time services. The algorithms considered are TOPSIS [8], ELECTRE [8][6] and GRA [16]. Amongst these MADM algorithms, GRA is found to be the
most suited for optimization objectives requiring both monotonic and non-monotonic utilities of attributes. Using a heterogeneous wireless networking environment as an example, we demonstrate how the algorithm can be implemented to achieve different optimization objectives, and its impacts on the resulting network ranking for network selection.

4.2.3 Comparison of MADM Algorithms for Use with Non-monotonic Utilities of Attributes

TOPSIS [8] calculates the perceived positive and negative ideal solutions based on the range of attribute values available for the alternatives, and selects the best solution as the one with the shortest distance to the positive ideal solution and longest distance from the negative ideal solution. The distances are measured in Euclidean terms. Because of the concept of positive and negative ideal solutions based on Euclidean distances, a standard implementation of TOPSIS requires that the utility of each attribute under consideration increases or decreases monotonically.

ELECTRE [6][8] performs pairwise comparisons among all alternatives for each one of the attributes separately in order to find outranking relationships between the alternatives. In its standard implementation, the method removes the less desirable alternatives. Using a complementary analysis it is possible to select the best suited alternatives. Since the comparison is directed amongst the available alternatives there is no concept in ELECTRE of comparing the alternatives to some reference set of values to see how close the parameters values are to the desired values. The notion of a monotonically increasing or decreasing utility of an attribute is inherent in a direct comparison amongst the alternatives, which makes the standard ELECTRE algorithm not well suited for use with attributes having non-monotonic utilities.

GRA is another very popular decision-making technique that is based on the Grey System Theory. Originally developed by Deng [16], the Grey Systems Theory has been applied to solve a variety of real life problems in many fields ranging from business, operations research, and engineering, to social sciences. One of its areas of application has been in MADM, where multiple attributes influence the decision process. Unlike MADM algorithms described earlier, GRA uses a reference set of attribute values for comparison with attribute values of the alternatives. It has been applied in the past [15][13][14] to solve the problem of network selection in a heterogeneous networking environment. The problem of selecting an optimal network in a heterogeneous networking environment, however, is quite complex and it is possible to apply GRA differently to the problem. For example, the utility aspects of the algorithm were not
explored in the prior work and a single reference network was constructed that implied a monotonic utility for all the attributes for all service or user types. Because of its ability to use reference attribute values in the decision process, GRA can be applied where the optimization objectives require a non-monotonic utility for some of the attributes and monotonic utility for the others. As described in the later section, such an implementation of GRA would use multiple reference networks. The network rankings in this case could be quite different than if a monotonic utility was considered for all the attributes.

It is clear that for optimization scenarios where a utility does not increase or decrease monotonically with an increase or decrease in attribute value, standard implementations of MADM algorithms such as TOPSIS and ELECTRE will have limited applicability. Other simpler compensating MADM algorithms such as SAW (Simple Additive Weighing) [5] and WPM (Weighed Product Method) [5] also have similar limitations because of their inherent assumption about monotonic utilities of attributes. These MADM algorithms do, however, allow assignment of different weights to the attributes before they are combined to calculate the ranking indices. This general feature of MADM algorithms can be used to apply algorithms such as TOPSIS to network selection for different service types as described in the previous chapter. For example, services that are less sensitive to QoS and more sensitive to the transport cost (e.g., web browsing), could have higher weights assigned to the cost attribute and lower weights assigned to the QoS attribute. This would allow alternatives that are closer to the positive ideal solution in the transport cost (i.e., lowest in cost) to receive more importance in decision making than network alternatives with QoS related attribute values closest to the positive ideal solution (i.e., best QoS attribute value). However, it is important to note that this type of applicability would provide a different optimization than trying to find the network alternative that is closest in QoS attributes to the requested service's QoS attributes. As stated earlier, GRA favors a selection that gives a closest match to a set of reference data values. This process inherently supports the notion that these reference values do not necessarily need to be the best or the worst values associated with the attributes. In addition, it also has the ability to assign different weights to different attributes. These two tunable aspects of GRA when combined provide a much better mechanism to achieve optimization objectives involving attributes with non-monotonic utilities. The rest of this section describes the application of GRA to the problem of network selection with some attributes having non-monotonic utilities.
4.2.4 Theory of Grey Relational Space

GRA is based on the concept of Grey Relational Space (GRS). GRS \((X,Y)\) describes a relationship \(Y\) between reference data values \(X_0\) and sequence of data values \(X\). So if \(y \in Y\), \(x_i \in X\), \(x_0 \in X_0\) such that \(x_0 = x_0(1), \ldots, x_0(n)\) and \(x_i = x_i(1), \ldots, x_i(n)\) then \(y(x_0(k),x_i(k))\) would represent a GRS at point \(k\) provided the axioms documented in [16] are satisfied. In addition, a Grey Relational Grade for a series \(i\) could then be represented as

\[
y(x_0, x_i) = \frac{1}{n} \sum_{k=1}^{n} y(x_0(k), x_i(k)).
\]

A representation of \(y(x_0(k), x_i(k))\) that satisfies all of the axioms in [8] is represented as:

\[
y(x_0(k), x_i(k)) = \frac{\min_{k} \min \left| x_0(k) - x_i(k) \right| + \xi \max_{k} \max \left| x_0(k) - x_i(k) \right|}{\left| x_0(k) - x_i(k) \right| + \xi \max_{k} \max \left| x_0(k) - x_i(k) \right|}
\]

where \(\xi\) is called a distinguished coefficient and has values ranging between zero and one. \(y(x_0(k), x_i(k))\) is called the Grey Relational Coefficient (GRC). When applying GRA to ranking of networks while selecting a network, GRC is a measure of how closely a network’s attributes match the reference network’s attributes. In this respect it represents the overall utility that takes into consideration individual values of all the attributes. The higher the value of GRC, the closer would be the network to the reference network. Hence, for the purposes of network selection the GRC is equivalence of a utility that takes into consideration all individual attribute values.

4.2.5 Application of GRA Adapted to Network Selection with Non-monotonic Utility

For the network selection example described here, the same set of attributes have been used as provided in section 4.1 and described in [4]. Using these attributes, a candidate network \(NW\) for evaluation by GRA can be represented by the follow attribute vector,

\[
NW = [CB \ TB \ AB \ U \ D \ J \ L]
\]

If there are \(N\) alternative networks to be considered in the selection process, they can be represented in the form of a matrix of network attributes as follows,
A reference access network is needed for application of GRA. In the case of monotonically increasing or decreasing utilities for the attributes, this reference network can be developed by using the maximum or minimum value of the attributes. In this case there will be only one reference network. However if there are services that have different QoS requirements (VoIP, streaming, web browsing etc.), or the user are of different categories (bronze, silver, gold) then the decision maker can use a different reference network for each one of the categories. A reference network in this case can be created based on the information about the user / terminal preferences, e.g., indication of requested service, and based on the user profile in the home network, e.g., the subscribed QoS. These multiple reference networks would result in non-monotonic utility for some of the attributes. Table 4-3 shows four different reference networks used in the example. The reference network i for a particular service or user type can therefore be represented as follows.

\[
(NW_{\text{ref}})_i = ((CB_{\text{ref}})_i, (TB_{\text{ref}})_i, (AB_{\text{ref}})_i, (U_{\text{ref}})_i, (D_{\text{ref}})_i, (J_{\text{ref}})_i, (L_{\text{ref}})_i)
\]

The units of measurement for the attributes such as cost, bandwidth and delay will be different. In order apply the algorithm without having the artifacts related to different units of measurement impacting the results, the attributes will have to be made unit-less before they can be directly compared or combined during the calculations. This process is called normalization [8]. Using these normalized attribute values, an updated matrix is created as follows,

\[
NW_i = \begin{bmatrix}
CB_1 & TB_1 & AB_1 & U_1 & D_1 & J_1 & L_1 \\
CB_2 & TB_2 & AB_2 & U_2 & D_2 & J_2 & L_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\end{bmatrix}
\]
The reference network i's attributes are also normalized and a normalized reference network vector is created as follows

\[(NW_{ref})_i = ((C_{B_{ref}})_i, (T_{B_{ref}})_i, (A_{B_{ref}})_i, (U_{ref})_i, (D_{ref})_i, (J_{ref})_i, (L_{ref})_i)\]

If the reference attribute values lie outside of the attributes values for the alternatives under considerations, then calculation of the maximum and minimum values to be used in the normalization process should include the reference values as well.

Distance vectors are calculated for attributes of each access network under consideration by taking the absolute difference between reference network attribute and the candidate network attribute. For example, in the case of TB for network i, the distance value from reference network j is calculated as follows

\[d_{TB} = |(TB_{ref})_i - TB_i|\]

The matrix of distance value for each of the attributes for the N networks under consideration can therefore be created as follows,

\[d_{NN} = \begin{bmatrix}
(d_{CB})_1 & (d_{TB})_1 & (d_{AB})_1 & (d_{U})_1 & (d_{D})_1 & (d_{J})_1 & (d_{L})_1 \\
(d_{CB})_2 & (d_{TB})_2 & (d_{AB})_2 & (d_{U})_2 & (d_{D})_2 & (d_{J})_2 & (d_{L})_2 \\
& & & & & & \\
& & & & & & \\
& & & & & & \\
& & & & & & \\
& & & & & & \\
& & & & & & \\
(d_{CB})_N & (d_{TB})_N & (d_{AB})_N & (d_{U})_N & (d_{D})_N & (d_{J})_N & (d_{L})_N \\
\end{bmatrix}\]

The GRC is a measure of the similarity of an attribute to its reference value. It is calculated for each of the matrix entries. For example, in the case of TB, it is calculated as follows

\[(GRC_{TB})_i = \frac{\Delta_{min} + \zeta \Delta_{max}}{\Delta_{TB} + \zeta \Delta_{max}}\]

where \(\zeta \in [0,1]\), and \(\Delta_{min}\) and \(\Delta_{max}\) can be calculated as follows

\[\Delta_{max} = \max_i (\Delta CB_i + \Delta TB_i + \Delta AB_i + \Delta U_i + \Delta D_i + \Delta J_i + \Delta L_i)\]

\[\Delta_{min} = \min_i (\Delta CB_i + \Delta TB_i + \Delta AB_i + \Delta U_i + \Delta D_i + \Delta J_i + \Delta L_i)\]

The next step is to consider the relative importance of each of the attributes in the decision about network selection. For this purpose each of the attribute is assigned a weight "w" such that
The weighted GRC matrix is obtained as follows

\[
GRC = \begin{bmatrix}
    w_{CB}^* (GRC_{CB}) & w_{TB}^* (GRC_{TB}) & w_{AB}^* (GRC_{AB})_1 & w_{D}^* (GRC_{D})_1 & w_{L}^* (GRC_{L})_1 \\
    w_{CB}^* (GRC_{CB})_2 & w_{TB}^* (GRC_{TB})_2 & w_{AB}^* (GRC_{AB})_2 & w_{D}^* (GRC_{D})_2 & w_{L}^* (GRC_{L})_2 \\
    \vdots & \vdots & \vdots & \vdots & \vdots \\
    w_{CB}^* (GRC_{CB})_N & w_{TB}^* (GRC_{TB})_N & w_{AB}^* (GRC_{AB})_N & w_{D}^* (GRC_{D})_N & w_{L}^* (GRC_{L})_N
\end{bmatrix}
\]

Using the GRC matrix thus calculated, the Grey Coefficient for each of the candidate network is calculated as follows.

\[
(GRC_{NW})_i = w_{CB}^* (GRC_{CB}) + w_{TB}^* (GRC_{TB}) + w_{AB}^* (GRC_{AB})_1 + w_{D}^* (GRC_{D})_1 + w_{L}^* (GRC_{L})_1
\]

The network with the highest value of Grey Coefficient is considered to be the best network.

### 4.2.6 Evaluation of Using Non-monotonic Utility in a Heterogeneous Wireless Network Environment

In order to evaluate the impact of different optimization objectives we consider a network selection scenario with five networks. For each of these networks, the attribute values to be used in the decision process are shown in Table 4-1. These attribute values represent a snapshot of network related information at the time of decision. The values in Table 4-1 are for illustrative purposes and are representative for listed example network types that a typical user could expect. For example, UMTS and 4G networks run on licensed spectrum and therefore their relative cost per byte is considered higher than that for unlicensed spectrum. The cost is also shown relatively higher for technologies with less efficient physical layer coding schemes (e.g., 801.11b vs 802.11n). Similarly the total bandwidth represents the estimated maximum throughput of a particular technology. In our case we address the network selection problem for three distinct types of services, namely VoIP, streaming media and web browsing. Each of these service types has its distinct set of QoS requirements. In the following subsection we describe how to use an adapted version of GRA for the scenarios under consideration.
4.2.6.1 Setting up GRA for Network Selection

The GRA algorithm can be applied in more than one ways to the problem of network selection. However it is well suited to handle diverse optimization objectives including those requiring non-monotonic utility of attributes. Figure 4.6 shows three different ways of application of the GRA algorithm. We recommend the third approach as described earlier as it provides the maximum flexibility for tuning the algorithm to different optimization objectives. A two step process for tuning GRA is proposed below.

Determine reference attribute values for different service or user categories

Based on the optimization criteria derived from the QoS requirements for the services or user types, reference networks are created by the decision maker (e.g., the user’s HN) It will be towards these reference values that the GRA will try to find a closest match from a given list of alternative networks. This step relates to addressing the non-monotonic nature of the utility for some of the attributes under consideration. For example, the VoIP reference network’s attribute values would reflect higher QoS requirements compared to a reference network for web browsing. Creating reference networks is a one time event and can therefore be provisioned into the decision process.

Figure 4.6 Three ways of using reference attribute values and attribute weights with GRA algorithm. Approach 3 is recommended in network selection.
Determine attribute weights for different service or user categories

To allow further tuning of the GRA algorithm to the optimization objectives, the weights (i.e., the importance) assigned to the attributes used in decision making are adjusted for each type of service. The weight assigned reflects the relative importance of an attribute for that service or user type. For example, the cost attribute would carry a relatively higher weight for streaming type service when compared with VoIP service. The process of determining attribute weights is also a one time event and can be provisioned into the decision process.

<table>
<thead>
<tr>
<th>Table 4-3 Reference attribute values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CB (%)</strong></td>
</tr>
<tr>
<td>Best QoS</td>
</tr>
<tr>
<td>VoIP</td>
</tr>
<tr>
<td>Streaming</td>
</tr>
<tr>
<td>Web Browsing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4-4 Assignment of attribute weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute weights used for scenarios 1, 3, 4 and 6</td>
</tr>
<tr>
<td><strong>CB</strong></td>
</tr>
<tr>
<td>VoIP</td>
</tr>
<tr>
<td>Streaming</td>
</tr>
<tr>
<td>Web Browsing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute weights used for scenarios 2 and 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CB</strong></td>
</tr>
<tr>
<td>VoIP, Streaming, Web Browsing</td>
</tr>
</tbody>
</table>

Table 4-3 shows four reference networks that were used in the evaluation. The values for CB, TB, AB and U attributes for the network are considered the type of attributes that the operator in
this scenario would like to optimize for the best value available amongst the alternatives. For example, by adopting the utilization of the lowest utilized alternative network as a reference attribute value, the decision maker could help improve the utilization for under utilized alternative networks and hence have an improved balance of traffic loads across different networks. However, this may not be always the desired optimization criterion and other decision makers may like to have different criteria for selecting this and other attribute values. So the first reference network is created from the best values for the attributes from amongst the alternatives networks. The remaining three reference networks use a combination of the best attribute values (for CB, TB and U) and reference values derived from QoS requirements for the service types (for AB, D, J and L). The entries in Table 4-3 show that some of the reference values generated from specific service types lie outside the range of attribute values for the network alternatives under consideration. This can potentially cause a problem in the normalization process. In order to avoid possible ranking abnormalities, the normalization process is modified by appropriately adjusting the minimum / maximum values used in order to include reference values as well.

The assigned weight for each of the attributes for different service categories considered in the evaluation is shown in Table 4-4. For example, the importance of the transport cost was considered high for web browsing as compared to VoIP. To evaluate the impacts of the assigned weights, a set of scenarios was also evaluated where only one weight distribution was used for all the different service types.

Scenario 1 uses the best values as the reference for all attributes but different weights for different service types. This is shown in the first approach of Figure 4.6. Scenario 2 only changes the reference attribute values while keeping the same attribute weights for all service types. This is shown in the second approach of Figure 4.6. Scenario 3 calculates ranking when the reference attribute values as well as the attribute weights were changed for different services as shown in approach 3 of Figure 4.6. For example, in the case of streaming media type service, it first compares the network attribute values with the reference values specific to streaming media type service to find the degree of match for each of the individual attributes. Then based on the emphasis that should be placed on the degree of match for each of the attributes, a weight is assigned to it as explained earlier.
Figure 4.7 Results for network detection for three possible configurations of GRA. Configuration 3 is preferred as it provides maximum flexibility to fine tuning the algorithm to optimization objectives.

Figure 4.7 shows the results for all the scenarios that were evaluated and hence shows the impact of using multiple reference networks and attribute weights. As explained earlier it is not possible to directly evaluate results of MADM algorithm in terms of accuracy. However since scenario 3 uses a combination of reference values and attribute weights to distinguish amongst different service types, it will provide a more balanced approach with results closest to optimization objectives of the decision maker if the intent was to select the network closest in characteristics to the reference network for that service type. By comparing results for scenarios 1 and 3 it is also apparent that using different reference values for different service types actually does impact the ranking. Similarly a comparison of scenarios 1 and 2 shows that the network rankings are
impacted if different distribution of attribute weights are used for each service type as opposed to a single distribution of attribute weights for all services.

4.3 Application of ELECTRE to Network Selection in Heterogeneous Wireless Access

The ability to provide a good and consistent customer experience in QoS demanding applications such as (VoIP or streaming media would depend upon the ability to select the most optimal delivery network. The decision of network selection is influenced by the optimization objectives of the decision maker. Selection of a non-optimal network can result in problems such as the unnecessary use of expensive access types or poor service experience.

4.3.1 Evaluating MADM Algorithms for Use in Network Selection

The decision process for the selection of a service delivery network takes into consideration several factors related to, e.g., access network capabilities, the current network conditions and transportation costs. Researchers have considered the use of MADM algorithms to rank the candidate networks in a preference order. There are numerous types of MADM algorithms with [5] documenting thirteen of them. Several alternate MADM algorithms can be suitable for solving a decision problem and the decision maker in this situation can be faced with the task of selecting the most appropriate method from amongst a number of feasible methods. Classification of MADM algorithms into categories [5] can help to eliminate the algorithms in categories that are not well suited to the problem space, but this process does not provide the most suited algorithm. It is conceivable that a suitable MADM algorithm may be selected for a particular decision problem based on one or both of the following criteria.

4.3.1.1 Accuracy of the results obtained from an algorithm

For a variety of reasons different algorithms, when applied to the same problem under the same assumptions, can result in different rankings of the alternatives. In such scenarios it is not possible to objectively rank the MADM algorithms for their ranking accuracy as it would require the use of another MADM algorithm to get such a ranking. For this reason it has been found difficult to use the accuracy of the results as a criterion in selecting a specific type of MADM algorithm.
4.3.1.2 Appropriateness of applying the algorithm to the problem

Because of differences in the approaches used by different MADM algorithms, a direct comparison amongst them is difficult. It has been proposed in the past that a method which is capable of solving the decision problem and whose decision making philosophy reflects the values of the decision maker can be considered to be the best suited. Decision makers in general prefer deterministic algorithms that provide reliable results based on a simple and easy to understand philosophy.

Here we describe the use of ELECTRE [5][6], a type of MADM algorithm, to the problem of network selection. ELECTRE algorithms perform a pair-wise comparison amongst the alternatives using each of the attributes under consideration, an approach that is very popular with decision makers because of its deterministic nature and simple philosophy. Other MADM algorithms such as TOPSIS [5][15] and SAW [5][15] have different decision making strategies but share the trait of simplicity in their philosophy with ELECTRE. Compared with these MADM algorithms, GRA algorithms [12][13][15] are more recent and the philosophy behind them are less intuitive and more complex. It is based on the Grey Systems Theory, which can best be compared to fuzzy mathematics and probabilistic decision making approaches. GRA provides a measure of similarity of a set of values to a set of reference values. The concept of reference values, as will be discussed in a later section, is very useful in network selection. Other MADM algorithms, such as ELECTRE, TOPSIS and SAW, do not have this capability as their comparison processes assume a monotonically increasing or decreasing utility or level of importance associated with each attribute value. Therefore despite the much more abstract decision philosophy of GRA, it has been applied to the problem of network selection.

The standard ELECTRE algorithm as indicated above has some shortcomings that if properly addressed would make it very attractive for application to the problem of network selection because of its simple decision making philosophy. ELECTRE assumes a monotonic utility and does not provide a complete ranking of all the alternatives either, which would be needed to find the top ranking candidate network. In this section an alternative approach to apply the ELECTRE algorithm has been developed so that it now provides a complete ranking of the networks under consideration. The algorithm has also been modified to make it suitable for application to scenarios where the utility of some attributes is non-monotonic. This change allows the application of the algorithm in scenarios where the decision maker would like to optimize the network selection to select the alternative that has attributes closest to a reference set of attribute
values. For example, it may be desired by the decision maker to select for web browsing the network that is not the best alternative from a QoS or cost perspective, but has attributes closest to the reference values for a desired network for web browsing as perceived by the decision maker. The modifications to the ELECTRE algorithm described in the next section would allow such selections and make it very well suited for ranking candidate networks for network selection.

4.3.2 Application of Modified ELECTRE to Network Selection

ELECTRE was developed by Bernard Roy [6] in the 1960s as a practical decision making tool and has found vast applications in engineering decision making problems. The method performs pair-wise comparisons among alternatives for each one of the attributes separately to establish outranking relationships between the alternatives [5][17]. In order to formulate network selection as a MADM problem the factors impacting the decision process have to be determined. The attributes used in the network selection decision process here are the same as provided in Section 4.1 and described in [4]. Using these attributes, from a decision making perspective the attributes of the i-th candidate network can be represented by a vector as follows,

\[ NW_i = [CB_i, TB_i, AB_i, U_i, D_i, J_i, L_i] \]

For N alternative networks to be considered in the selection process, a matrix of network attributes can be formulated as follows,

\[
NW = \begin{bmatrix}
CB_1 & TB_1 & AB_1 & U_1 & D_1 & J_1 & L_1 \\
CB_2 & TB_2 & AB_2 & U_2 & D_2 & J_2 & L_2 \\
& & & & & & \\
& & & & & & \\
& & & & & & \\
CB_N & TB_N & AB_N & U_N & D_N & J_N & L_N
\end{bmatrix}
\]

In order to best match the optimization objectives described in the previous section, we modify the basic ELECTRE method by utilizing a reference network, which can be considered to be an access network that has a desired set of attribute values given in the reference attribute vector. This reference attribute vector is used to adjust the raw attribute values for the alternative networks before they are compared. The source of reference attribute values can be different depending upon the decision process used. For example, it is possible that the information is provided by the user terminal via indication of the service it wants to initiate. In another scenario
the information can be provided by the user’s HN operator via its knowledge about the subscribed QoS in the user profile. We can represent this reference access network as,

\[ \text{NW}_{\text{ref}} = (\text{CB}_{\text{ref}}, \text{TB}_{\text{ref}}, \text{AB}_{\text{ref}}, \text{U}_{\text{ref}}, \text{D}_{\text{ref}}, \text{J}_{\text{ref}}, \text{L}_{\text{ref}}) \]

In the proposed modification to the algorithm, the value of each of the attributes in matrix NW is compared with a corresponding reference attribute value. An absolute difference between the two values is taken to calculate a new matrix as follows. In standard ELECTRE algorithm this step would be skipped with the assumption of a monotonically increasing or decreasing utility for the raw attribute value.

\[
\text{NW}_{\text{Adjusted}} = \begin{bmatrix}
|CB_1 - CB_{\text{ref}}| & |TB_1 - TB_{\text{ref}}| & |AB_1 - AB_{\text{ref}}| & |U_1 - U_{\text{ref}}| & |D_1 - D_{\text{ref}}| & |J_1 - J_{\text{ref}}| & |L_1 - L_{\text{ref}}| \\
|CB_2 - CB_{\text{ref}}| & |TB_2 - TB_{\text{ref}}| & |AB_2 - AB_{\text{ref}}| & |U_2 - U_{\text{ref}}| & |D_2 - D_{\text{ref}}| & |J_2 - J_{\text{ref}}| & |L_2 - L_{\text{ref}}| \\
& & & & & & \\
& & & & & & \\
& & & & & & \\
|CB_N - CB_{\text{ref}}| & |TB_N - TB_{\text{ref}}| & |AB_N - AB_{\text{ref}}| & |U_N - U_{\text{ref}}| & |D_N - D_{\text{ref}}| & |J_N - J_{\text{ref}}| & |L_N - L_{\text{ref}}| 
\end{bmatrix}
\]

In order to remove the impact of use of different measurement units (e.g., dollars/byte for transportation cost vs. milliseconds for latency or jitter) the attributes represented in the matrix have to be normalized. Since in the proposed modification to the ELECTRE algorithm, the raw attribute values have been adjusted with respect to the reference attribute values, it can now be assumed that for the adjusted values, the larger the attribute value, the farther it is from the desired or the reference value. In other words, all attribute values can now be considered to have a monotonically decreasing utility. Since a lower value for an adjusted attribute is considered an indication of a better network in the selection process, each attribute \( X_j \) in row \( i \) of a specific column of the matrix can be normalized as follows,

\[
\tilde{X}_i = \frac{\max_{j=1..N} (X_j) - X_i}{\max_{j=1..N} (X_j) - \min_{j=1..N} (X_j)}
\]

A normalized matrix with these normalized values as its elements is created as follows.
During the process of overall comparison of the alternatives, the impact of pair-wise comparison of different attributes is summed up. This summation should take into consideration the relative importance of each of the attributes involved in the decision about network selection. The information about the relative importance of the attributes can have similar sources as described earlier in generating a reference attribute vector. For example, a user may request the use of VoIP service whereby the relative importance of the transportation cost and total bandwidth is considered low because of VoIP being a low bit rate application. However factors such as latency, jitter are quite important for the VoIP type service. On the other hand the weight related information can also come from the user profile that can, e.g., indicate the user to be a Bronze user and hence assign a higher weight to the cost attribute and lower weights for the latency and jitter attributes. Therefore, depending upon the information about the service to be used or the user QoS profile, the j-th attribute is assigned a weight \( w_j \), such that

\[
W = w_{CB} + w_{TB} + w_{AB} + w_U + w_D + w_J + w_L = 1
\]

Using the assigned weights, an updated matrix is calculated as follows.

\[
\mathbf{NW}_{wl} = \begin{bmatrix}
w_{CB} * C_{\bar{B}}_1 & w_{TB} * T_{\bar{B}}_1 & w_{AB} * A_{\bar{B}}_1 & w_U * \bar{U}_1 & w_D * \bar{D}_1 & w_J * \bar{J}_1 & w_L * \bar{L}_1 \\
w_{CB} * C_{\bar{B}}_2 & w_{TB} * T_{\bar{B}}_2 & w_{AB} * A_{\bar{B}}_2 & w_U * \bar{U}_2 & w_D * \bar{D}_2 & w_J * \bar{J}_2 & w_L * \bar{L}_2 \\
w_{CB} * C_{\bar{B}}_N & w_{TB} * T_{\bar{B}}_N & w_{AB} * A_{\bar{B}}_N & w_U * \bar{U}_N & w_D * \bar{D}_N & w_J * \bar{J}_N & w_L * \bar{L}_N
\end{bmatrix}
\]

In order to compare the network alternatives, the concept of **concordance** and **discordance** has been introduced in ELECTRE, which are measures of satisfaction and dissatisfaction of the decision maker when one alternative is compared with another. Thus concordance and discordance sets are calculated, where a concordance set (CSet) provides a list of attributes for which an alternative network under consideration is better than another alternative network it is
being compared with, and a discordance set (DSet) on the other hand provides a list of attributes where the alternative network under consideration is worse than the compared alternative. For example when network 1 is being compared with network 2, the concordance set $CSet_{12}$ is the subset of all attributes that indicate that network 1 should be preferred over network 2, and the discordance set $DSet_{12}$ is the subset of all attributes that indicate a preference of network 2 over network 1. Mathematically this can be represented as follows,

$$CSet_{12} = \{ j : (NW_{\text{norm}})_{1,j} >= (NW_{\text{norm}})_{2,j} \}$$

$$DSet_{12} = \{ j : (NW_{\text{norm}})_{1,j} < (NW_{\text{norm}})_{2,j} \}$$

Using the concordance and discordance sets, corresponding matrices are constructed. ELECTRE calculates the elements of concordance matrix $C$ as follows,

$$C_{kl} = \sum_{j \in CSet_{kl}} W_j$$

The concordance matrix $C$ can be represented as,

$$C = \begin{bmatrix}
- & C_{12} & \ldots & C_{1N} \\
C_{21} & - & \ldots & C_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
C_{N1} & C_{N2} & \ldots & -
\end{bmatrix}$$

The entries for the concordance matrix are not defined for the diagonal. Similarly, ELECTRE defines the elements of discordance matrix as follows,

$$d_{kl} = \frac{\sum_{j \in DSet_{kl}} |(NW_{\text{norm}})_{kj} - (NW_{\text{norm}})_{lj}|}{\sum_{i} |(NW_{\text{norm}})_{kj} - (NW_{\text{norm}})_{lj}|}$$

Similarly, the discordance matrix $D$ can be represented as,

$$D = \begin{bmatrix}
- & D_{12} & \ldots & D_{1N} \\
D_{21} & - & \ldots & D_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
D_{N1} & D_{N2} & \ldots & -
\end{bmatrix}$$
The entries for the discordance matrix are also not defined for the diagonal.

Two possible approaches can be considered to proceed further in decision making based on the ELECTRE algorithm.

4.3.2.1 Approach 1

In the standard ELECTRE method, the outranking calculations are performed as follows.

The concordance and discordance dominance matrices are determined. The concordance dominance matrix is calculated using a threshold value for the concordance index. A way to determine the threshold value, \( C_{\text{threshold}} \), is to use the average concordance index as follows,

\[
C_{\text{threshold}} = \frac{\sum_{k=1}^{N} \sum_{l=1}^{N} c_{kl}}{N \times (N - 1)}
\]

Using the \( C_{\text{threshold}} \) value, elements of the concordance dominance matrix, \( C_{\text{dom}} \), are calculated as follows,

\[
(C_{\text{dom}})_{kl} = \begin{cases} 1 : c_{kl} \geq C_{\text{threshold}} \\ 0 : c_{kl} < C_{\text{threshold}} \end{cases}
\]

The discordance dominance matrix is calculated using a similar threshold value, \( D_{\text{threshold}} \). This value can be calculated using a similar formula as follows,

\[
D_{\text{threshold}} = \frac{\sum_{k=1}^{N} \sum_{l=1}^{N} d_{kl}}{N \times (N - 1)}
\]

Using the \( D_{\text{threshold}} \) value, elements of the discordance dominance matrix, \( D_{\text{dom}} \), are calculated as follows,

\[
(D_{\text{dom}})_{kl} = \begin{cases} 0 : d_{kl} \geq D_{\text{threshold}} \\ 1 : d_{kl} < D_{\text{threshold}} \end{cases}
\]

The aggregate dominance matrix, \( A_{\text{dom}} \), is calculated as follows,

\[
(A_{\text{dom}})_{kl} = (C_{\text{dom}})_{kl} \times (D_{\text{dom}})_{kl}
\]
The aggregate dominance matrix is able to provide partial preference ordering of the access networks under consideration. For example if \((A_{\text{dom}})_{12} = 1\), then this would imply that network 1 is preferred over network 2 when both concordance and discordance criteria are used.

A problem with this approach in formulating the ELECTRE algorithm is the arbitrary selection of threshold values. These threshold values can significantly impact the outcome of the algorithm. In addition the results of this ELECTRE method do not provide a complete ranking for all the alternatives.

4.3.2.2 Approach 2

The complementary analysis in [8] tries to address the shortcomings of approach 1. A new parameter \(C_i\), called the net concordance index is calculated. \(C_i\) is a measure of the dominance of an alternative \(i\) over other alternatives when compared with a measure of dominance of other alternatives over the alternative \(i\). It can be calculated as follows,

\[
C_i = \sum_{j=1 \atop j \neq i}^{N} C_{ij} - \sum_{j=1 \atop j \neq i}^{N} C_{ji}
\]

Similarly, the term net discordance index \(D_i\), is defined as a measure of relative weakness of alternative \(i\) over other alternatives when compared with a measure of weakness of other alternatives from the alternative \(i\).

\[
D_i = \sum_{j=1 \atop j \neq i}^{N} D_{ij} - \sum_{j=1 \atop j \neq i}^{N} D_{ji}
\]

An alternative with the highest value of net concordance index \(C\) and lowest value of net discordance index \(D\) would be preferred. It is possible that the alternative with the highest value of concordance index is not the same as that with the lowest value of discordance index. In order to address this issue, the alternatives are ranked based on the concordance and discordance indices and each alternative is ranked by taking the average of these two rankings. The alternative with the highest average ranking is considered to be the best alternative. Alternatives with the same average ranking would be considered equally suited.

For the case of network selection, in order to find the top candidate network, it is required to have a complete ranking for all the networks under consideration. Therefore approach 2 has been applied as it provides a clearer ranking of the alternatives.
4.3.3 Evaluation of Modified ELECTRE

In order to evaluate the use of ELECTRE as well as the impact of the proposed changes to the algorithm, we consider a network selection situation with five network types to choose from. The attribute values for these networks, determined at the time of network selection, are provided in Table 4-5.

We consider the scenario where network selection is influenced by the requested service indicated by the user. Three services, namely VoIP (low bit rate, real-time), streaming (high bit rate, soft real-time), and web browsing (varying bit rate, bursty, non real-time) are considered. The service type is used to assign attribute weights. For example in the case of VoIP, since it is a low bandwidth application, the total bandwidth and available bandwidth are not considered important and therefore assigned a weight of zero. Also the transport cost is not considered significant because of the higher revenue generating nature of VoIP applications but attributes such as low latency and jitter are quite significant for a good customer experience and therefore assigned higher weights. The values of the assigned weights for different services considered are provided in Figure 4.8. For the case of the modified algorithm, different reference values for the attributes are used for each of the service type. These reference values indicate the preferred attribute values for the service type. Table 4-6 provides these values for VoIP, streaming and web browsing applications.

The results are documented in Tables 4-7 and 4-8. As described in alternative approach 2, the rankings shown in these tables are the averages of two rankings obtained using concordance and discordance indices. So the highest rankings in these tables may not actually be 1 unless there is a network which is the best both from the perspective of concordance and discordance indices. Table 4-7 shows the network rankings when a standard version of ELECTRE as described earlier was used with approach 2. The results show that the same network, i.e., #3 is being selected for all the services although the rankings for the rest of the networks are different for the services; e.g., the second ranked network for VoIP service is different than that for streaming or web browsing service. The rankings for networks using the modified version of the algorithm that uses reference attribute values (as shown in Table 4-6) is shown in Table 4-8. The selected network rankings in this case are different for VoIP and streaming. For web browsing, three networks are ranked at the same level. To further explain the reason for the selection of different networks by the algorithm Tables 4-9, 4-10 and 4-11 are provided. These tables show how the input attribute values are changed by the use of reference values, normalization and then use of attribute weights. It can be
seen from these tables that the adjusted, normalized and weighed attribute values that form the
input to the algorithm are quite different in the case of VoIP, streaming and web browsing
services. This reflects the effect of data manipulation performed to meet the optimization
objectives of the decision maker. As a result, different networks can get selected for different
service types.

Table 4-5 Attribute Values for Scenarios under Consideration

<table>
<thead>
<tr>
<th>Network</th>
<th>CB</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (ms)</th>
<th>J (ms)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nwk1 e.g. UMTS</td>
<td>100</td>
<td>2</td>
<td>0.2</td>
<td>10</td>
<td>400</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Nwk2 e.g. 802.11b</td>
<td>20</td>
<td>11</td>
<td>1</td>
<td>20</td>
<td>200</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Nwk3 e.g. 802.11a</td>
<td>10</td>
<td>54</td>
<td>2</td>
<td>20</td>
<td>100</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Nwk4 e.g. 802.11n</td>
<td>5</td>
<td>100</td>
<td>5</td>
<td>20</td>
<td>150</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Nwk5 e.g. 4G</td>
<td>30</td>
<td>100</td>
<td>5</td>
<td>20</td>
<td>100</td>
<td>20</td>
<td>15</td>
</tr>
</tbody>
</table>

Figure 4.8 Weights associated with attributes for different services

Table 4-6 Reference attribute values for voice over IP, streaming and web browsing services

<table>
<thead>
<tr>
<th>Service</th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (ms)</th>
<th>J (ms)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VoIP</td>
<td>5</td>
<td>100</td>
<td>0.02</td>
<td>10</td>
<td>100</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Streaming</td>
<td>5</td>
<td>100</td>
<td>0.1</td>
<td>10</td>
<td>400</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Web Browsing</td>
<td>5</td>
<td>100</td>
<td>0.1</td>
<td>10</td>
<td>1000</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 4-7 Ranking for networks using standard ELECTRE method with Alternative 2

<table>
<thead>
<tr>
<th>Network#1 e.g. UMTS</th>
<th>VoIP</th>
<th>Streaming</th>
<th>Web Browsing</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Network#2 e.g. 802.11b</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Network#3 e.g. 802.11a</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Network#4 e.g. 802.11n</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Network#5 e.g. 4G</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 4-8 Ranking for networks using modified ELECTRE method with Alternative 2

<table>
<thead>
<tr>
<th>Network#1 e.g. UMTS</th>
<th>VoIP</th>
<th>Streaming</th>
<th>Web Browsing</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Network#2 e.g. 802.11b</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Network#3 e.g. 802.11a</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Network#4 e.g. 802.11n</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Network#5 e.g. 4G</td>
<td>2</td>
<td>4.5</td>
<td>5</td>
</tr>
</tbody>
</table>
### Table 4-9 voice over IP service

#### Adjusted Attribute Values

<table>
<thead>
<tr>
<th></th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10⁶)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>95.00</td>
<td>98.00</td>
<td>0.18</td>
<td>0.00</td>
<td>300.00</td>
<td>35.00</td>
<td>85.00</td>
</tr>
<tr>
<td>#2</td>
<td>15.00</td>
<td>89.00</td>
<td>0.98</td>
<td>10.00</td>
<td>100.00</td>
<td>10.00</td>
<td>5.00</td>
</tr>
<tr>
<td>#3</td>
<td>5.00</td>
<td>46.00</td>
<td>1.98</td>
<td>10.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>#4</td>
<td>0.00</td>
<td>0.00</td>
<td>4.98</td>
<td>30.00</td>
<td>50.00</td>
<td>15.00</td>
<td>5.00</td>
</tr>
<tr>
<td>#5</td>
<td>25.00</td>
<td>0.00</td>
<td>4.98</td>
<td>10.00</td>
<td>0.00</td>
<td>5.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

#### Normalized Attribute Values

<table>
<thead>
<tr>
<th></th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10⁶)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>#2</td>
<td>0.842</td>
<td>0.092</td>
<td>0.833</td>
<td>0.667</td>
<td>0.667</td>
<td>0.714</td>
<td>0.941</td>
</tr>
<tr>
<td>#3</td>
<td>0.947</td>
<td>0.531</td>
<td>0.625</td>
<td>0.667</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>#4</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.833</td>
<td>0.571</td>
<td>0.941</td>
</tr>
<tr>
<td>#5</td>
<td>0.737</td>
<td>1.000</td>
<td>0.000</td>
<td>0.667</td>
<td>1.000</td>
<td>0.857</td>
<td>1.000</td>
</tr>
</tbody>
</table>

#### Normalized and Weighted Attribute Values

<table>
<thead>
<tr>
<th></th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10⁶)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.200</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>#2</td>
<td>0.042</td>
<td>0.000</td>
<td>0.000</td>
<td>0.133</td>
<td>0.200</td>
<td>0.214</td>
<td>0.141</td>
</tr>
<tr>
<td>#3</td>
<td>0.047</td>
<td>0.000</td>
<td>0.000</td>
<td>0.133</td>
<td>0.300</td>
<td>0.300</td>
<td>0.150</td>
</tr>
<tr>
<td>#4</td>
<td>0.050</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.250</td>
<td>0.171</td>
<td>0.141</td>
</tr>
<tr>
<td>#5</td>
<td>0.037</td>
<td>0.000</td>
<td>0.000</td>
<td>0.133</td>
<td>0.300</td>
<td>0.257</td>
<td>0.150</td>
</tr>
</tbody>
</table>
Table 4-10 Steaming service

**ADJUSTED ATTRIBUTE VALUES**

<table>
<thead>
<tr>
<th></th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>95.00</td>
<td>98.00</td>
<td>0.80</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>50.00</td>
</tr>
<tr>
<td>#2</td>
<td>15.00</td>
<td>89.00</td>
<td>0.00</td>
<td>10.00</td>
<td>200.00</td>
<td>25.00</td>
<td>30.00</td>
</tr>
<tr>
<td>#3</td>
<td>5.00</td>
<td>46.00</td>
<td>1.00</td>
<td>10.00</td>
<td>300.00</td>
<td>35.00</td>
<td>35.00</td>
</tr>
<tr>
<td>#4</td>
<td>0.00</td>
<td>0.00</td>
<td>4.00</td>
<td>30.00</td>
<td>250.00</td>
<td>20.00</td>
<td>30.00</td>
</tr>
<tr>
<td>#5</td>
<td>25.00</td>
<td>0.00</td>
<td>4.00</td>
<td>10.00</td>
<td>300.00</td>
<td>30.00</td>
<td>35.00</td>
</tr>
</tbody>
</table>

**NORMALIZED ATTRIBUTE VALUES**

<table>
<thead>
<tr>
<th></th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.800</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>#2</td>
<td>0.842</td>
<td>0.092</td>
<td>1.000</td>
<td>0.667</td>
<td>0.333</td>
<td>0.286</td>
<td>1.000</td>
</tr>
<tr>
<td>#3</td>
<td>0.947</td>
<td>0.531</td>
<td>0.750</td>
<td>0.667</td>
<td>0.000</td>
<td>0.000</td>
<td>0.750</td>
</tr>
<tr>
<td>#4</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.167</td>
<td>0.429</td>
<td>1.000</td>
</tr>
<tr>
<td>#5</td>
<td>0.737</td>
<td>1.000</td>
<td>0.000</td>
<td>0.667</td>
<td>0.000</td>
<td>0.143</td>
<td>0.750</td>
</tr>
</tbody>
</table>

**NORMALIZED AND WEIGHTED ATTRIBUTE VALUES**

<table>
<thead>
<tr>
<th></th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.160</td>
<td>0.200</td>
<td>0.100</td>
<td>0.100</td>
<td>0.000</td>
</tr>
<tr>
<td>#2</td>
<td>0.168</td>
<td>0.014</td>
<td>0.200</td>
<td>0.133</td>
<td>0.033</td>
<td>0.029</td>
<td>0.050</td>
</tr>
<tr>
<td>#3</td>
<td>0.190</td>
<td>0.080</td>
<td>0.150</td>
<td>0.133</td>
<td>0.000</td>
<td>0.000</td>
<td>0.038</td>
</tr>
<tr>
<td>#4</td>
<td>0.200</td>
<td>0.150</td>
<td>0.000</td>
<td>0.000</td>
<td>0.017</td>
<td>0.043</td>
<td>0.050</td>
</tr>
<tr>
<td>#5</td>
<td>0.147</td>
<td>0.150</td>
<td>0.000</td>
<td>0.133</td>
<td>0.000</td>
<td>0.014</td>
<td>0.038</td>
</tr>
</tbody>
</table>
Table 4-11 Web browsing service

**ADJUSTED ATTRIBUTE VALUES**

<table>
<thead>
<tr>
<th></th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>95.00</td>
<td>98.00</td>
<td>0.10</td>
<td>0.00</td>
<td>600.00</td>
<td>50.00</td>
<td>0.00</td>
</tr>
<tr>
<td>#2</td>
<td>15.00</td>
<td>89.00</td>
<td>0.90</td>
<td>10.00</td>
<td>800.00</td>
<td>75.00</td>
<td>80.00</td>
</tr>
<tr>
<td>#3</td>
<td>5.00</td>
<td>46.00</td>
<td>1.90</td>
<td>10.00</td>
<td>900.00</td>
<td>85.00</td>
<td>85.00</td>
</tr>
<tr>
<td>#4</td>
<td>0.00</td>
<td>0.00</td>
<td>4.90</td>
<td>30.00</td>
<td>850.00</td>
<td>70.00</td>
<td>80.00</td>
</tr>
<tr>
<td>#5</td>
<td>25.00</td>
<td>0.00</td>
<td>4.90</td>
<td>10.00</td>
<td>900.00</td>
<td>80.00</td>
<td>85.00</td>
</tr>
</tbody>
</table>

**NORMALIZED ATTRIBUTE VALUES**

<table>
<thead>
<tr>
<th></th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>#2</td>
<td>0.842</td>
<td>0.092</td>
<td>0.833</td>
<td>0.667</td>
<td>0.333</td>
<td>0.286</td>
<td>0.059</td>
</tr>
<tr>
<td>#3</td>
<td>0.947</td>
<td>0.531</td>
<td>0.625</td>
<td>0.667</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>#4</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.167</td>
<td>0.429</td>
<td>0.059</td>
</tr>
<tr>
<td>#5</td>
<td>0.737</td>
<td>1.000</td>
<td>0.000</td>
<td>0.667</td>
<td>0.000</td>
<td>0.143</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**NORMALIZED AND WEIGHED ATTRIBUTE VALUES**

<table>
<thead>
<tr>
<th></th>
<th>CB (%)</th>
<th>TB (Mbps)</th>
<th>AB (Mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.150</td>
<td>0.100</td>
<td>0.050</td>
<td>0.050</td>
<td>0.100</td>
</tr>
<tr>
<td>#2</td>
<td>0.421</td>
<td>0.005</td>
<td>0.125</td>
<td>0.067</td>
<td>0.017</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>#3</td>
<td>0.474</td>
<td>0.027</td>
<td>0.094</td>
<td>0.067</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>#4</td>
<td>0.500</td>
<td>0.050</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>0.021</td>
<td>0.006</td>
</tr>
<tr>
<td>#5</td>
<td>0.368</td>
<td>0.050</td>
<td>0.000</td>
<td>0.067</td>
<td>0.000</td>
<td>0.007</td>
<td>0.000</td>
</tr>
</tbody>
</table>
4.4 Conclusion

Convergence of handheld multimedia devices with communication devices has created a new class of feature rich consumer devices which can provide seamless network connectivity for service delivery through a variety of communication technologies. In the face of a multitude of service delivery network options and the presence of heterogeneous wireless networks, the selection of an optimal network for service delivery is a major issue. MADM algorithms are a popular decision making tool in operations research and have also been considered for application in other areas. They are however known to suffer from the problem of ranking abnormalities. These abnormalities can potentially decrease the quality of the results. We have applied TOPSIS, an MADM algorithm, to the problem of network selection and analyzed the causes of the ranking abnormalities problem for TOPSIS as applied to the problem of network selection. We have proposed an iterative approach for application of TOPSIS, which can improve the results for network selection by comparing only the more likely candidates in the process. Results for the iterative method have been presented to show that the top candidate network determined by TOPSIS can change when a candidate at the bottom of the ranking is removed from the comparison, and the iterative approach gives a more consistent and accurate final result. The network selection algorithm described in this chapter is well suited for use with service delivery architecture described in Chapter 2 [7].

Prior research on application of MADM algorithms to the problem of network selection has not compared them to select the most appropriate algorithm. We have discussed the decision maker’s optimization objectives and hence the utility of attributes as a means to evaluate the algorithms and select the most appropriate one. The need to support non-monotonic utility for attributes in order to handle diverse optimization objectives of a decision maker has been shown and MADM algorithms have been evaluated for handling these objectives. We have shown that many of the commonly used MADM algorithms such as SAW, WPM, TOPSIS, and ELECTRE in their standard form are not well suited to the network selection problem because of assumptions about monotonically increasing or decreasing utilities of the attributes. We have also shown that GRA can easily be adapted to use multiple reference networks and is therefore better suited for achieving this type of optimization objectives. The evaluation of adapted GRA in this chapter has also demonstrated that the selection of delivery network will be impacted by how the algorithm is used to achieve the optimization objectives. A novel two step process has been proposed that uses multiple reference values. It has been explained through an example that shows how reference
attribute values and attribute weights impact the selection process. The adjustment of these parameters for different service types has been discussed. The decision process proposed here can be used in a heterogeneous wireless network system environment.

We have described the adaptation of ELECTRE, an MADM algorithm, for ranking network alternatives during the network selection process. The use of a particular MADM algorithm for a specific problem is based on an assessment about the appropriateness of the algorithm for application to the problem space. Here we describe the use of a modified algorithm that provides a complete ranking of alternative networks. The modifications also allow the usage of ELECTRE with attributes exhibiting a non-monotonic utility. These modifications make the algorithm more adept to application in network selection as it expands its applicability to a wider range of optimization objectives. Such network selection scenarios are of special importance in a heterogeneous wireless networking environment being used to deliver a variety of service types. With these modifications and a simple decision making philosophy, ELECTRE is an ideal algorithm for use in network selection. In order to evaluate the proposed use of the algorithm, an example has been described where depending upon the QoS requirements of the service being requested by the user device, a different delivery network maybe chosen. The results have been compared with and without implementing the modifications to support non-monotonic utility of attributes. The proposed algorithm can be used with the network selection architecture described in Chapter 2 [7].
4.5 References


5 Application of Data Prediction and Fuzzy Techniques with MADM-Based Automated Network Selection

5.1 Introduction

Broadband wireless networks such as WLAN and WWAN have one characteristic in common in that they all use IP as the data transport. In order to provide ubiquitous coverage, increasingly these all-IP wireless technologies are being made to interwork. This makes the achievement of a consistent service experience over heterogeneous wireless technologies very important. Unlike a circuit-switched service environment, packet-switched IP networks are known to vary in terms of QoS. The variation from a network to network can be static such as because of the inherently different capabilities of the network (e.g., IEEE 802.11a vs. IEEE 802.11n) or it can be dynamic based on the network’s current congestion level. Selection of an optimal service delivery AN is an important problem to be solved in such an environment. Since a number of network attributes, e.g., those related to QoS have an impact on such decisions, the use of MADM algorithms [1][2] has been proposed in the past for ranking candidate ANs in terms of their suitability. Output of these algorithms is dependent upon the accuracy of information being used as input to them. In real world scenarios, often some of the information is either not available or it is imprecise. In addition in some cases, because of the candidate network types the usage of fuzzy input information is more useful than crisp values. Figure 5.1 illustrates that even when all the

---

1 Parts of this chapter have been published in the following
information is not available it can still be possible to narrow down the list of candidate networks. Below we describe scenarios where application of MADM algorithms for network selection while relying on incomplete information can be beneficial.

![Diagram](image)

**Figure 5.1** Use of MADM with imprecise information can help narrow down the candidate service delivery network options

### 5.2 Using Non-Compensating MADM Algorithms with Incomplete information

In the network selection process described in Chapter 3, initially, non-compensating MADM algorithms are used to narrow down the candidate list. The information used in this process comes from both the terminal and the entities within the network. However in some cases, the terminal or entities within the network may be either unable to or unwilling to provide some of the information to the decision making entity. For example, the information related to the terminal’s mobility profile may not be available for the decision making entity because the user / terminal has not indicated it. In other cases the network may not have all the information about the candidate networks that is to be used in the non-compensatory part of the MADM algorithm. For example, it may not know the exact coverage area or the authentication methods supported by the roaming partner’s ANs. Even in such scenarios by using the mechanisms described in Chapter 3 candidate networks can be narrowed down to fewer and more probable service delivery networks. The candidate networks in the resultant shortened list in such scenarios may not all be accessible to the terminal but the list can provide guidance to the terminal to help it narrow down the service delivery options.
In general, the attributes used in the non-compensating part of the algorithm can be separated into attributes essential for the decision process and those that can be considered less critical. The decision maker can decide which attributes it considers absolutely essential to get useable information from the decision process. For example, without the information about the location of the terminal, it is not possible to narrow down the search whereas service to be used can be considered less critical. Also, in such scenarios there will be a possibility that some of the top ranking networks in the short list are not accessible to the terminal; e.g., if the authentication mechanism related information was lacking for the network or terminal type. Therefore, to make the information useful, a list of preferred networks instead of a single top ranking network should be provided to the terminal.

5.3 Using Compensating MADM Algorithms with Data Prediction and Fuzzy Input

The application of a standard compensating MADM algorithm to network selection involves

- Identifying all alternatives and compensatory MADM attributes impacting the decision process,
- Assigning relative importance in the decision making process to each of the attributes, and
- Using a MADM algorithm to get a ranking of the alternatives.

These algorithms are used to determine the ranking of alternatives in terms of their desirability with respect to multiple criteria as a whole that can influence the decision. For the compensating MADM algorithm application, the following two scenarios have been identified where fuzzy approaches will be useful.

5.3.1 Scenario 1 - Imprecise or Missing Information

For compensating MADM algorithms to work properly, the attribute values have to be reliable in order to select the optimal service delivery network. However, as shown in Figure 5.2, due to the geographic distribution of the locations of the data collection points [3] it may not be possible to get real time updates for some of the attributes used in network selection. Also, while dealing with heterogeneous access technologies spanning autonomous operator domains, it may not be possible to get a homogeneous set of attribute data spread evenly over time that would allow a direct comparison between the networks towards their suitability for delivering the requested services.
In many cases there are measurement errors associated with the monitoring and processing of dynamic QoS attribute values such as packet delay, jitter, and loss. Hence there is a need to develop a mechanism of network selection when input attribute values are less reliable or unavailable.

Lotfi Zadeh [4][5] developed in the 1960s the theory of fuzzy logic that enables algebraic manipulation of fuzzy or imprecise data. Since then fuzzy logic theory has found applications in a variety of areas including decision support. Fuzzy numbers have been used in the past with MADM algorithms for scenarios where input attribute values are imprecise or hard to calculate [6][7]. The use of a very simple fuzzy logic based multiple-criteria decision-making system has been proposed [8] to perform vertical handovers in a heterogeneous networking environment. The application of this work to network selection is therefore limited. In [9] fuzzy based network selection has been discussed within the context of peer-to-peer networking. In [10] and [11] fuzzy mechanisms have been considered within the context of vertical handoffs. [10] actually converts fuzzy data to crisp values before applying standard MADM algorithms. [11] uses a fuzzy inference engine along with neural nets for predictions about the number of users. The work in [9]-[11] however has a somewhat different focus and does not take into consideration important...
practical aspects for the problem under consideration in our research, such as prediction of unavailable data, selection of fuzzy number type to represent predicted attributes, suitable defuzzification techniques or even a fundamental consideration about when it is appropriate to use a fuzzy MADM algorithm.

In this chapter we describe both types of scenarios where fuzzy techniques will be helpful and where other mechanisms are better suited. We propose a novel way of combining fuzzy techniques along with parameter estimation in network selection. They have been applied to the decision process while using a proposed fuzzy implementation of GRA, a MADM algorithm. We describe mechanisms for prediction/estimation of missing data, fuzzification of the estimated values and the subsequent defuzzification of network rankings obtained by the use of fuzzy GRA. Additional decision support tools that can be applied under such conditions are presented to obtain the Confidence Levels either in the network rankings or in the data before network rankings are obtained. Both approaches are described in this chapter. Together, the techniques described here improve the reliability of the results and allow decisions to be made under uncertain conditions. The mechanisms described in this chapter would work well with the network-assisted terminal-based network selection architecture described in Chapter 2 [3].

Figure 5.3 Steps involved in proposed network selection mechanism with imprecise information
Figure 5.3 shows a comprehensive approach towards a network selection mechanism that leverages parameter estimation techniques, fuzzy theory, MADM algorithms and also introduces the new concept of Confidence Level. The first step is to check if the service being requested or the user's subscription profile is sensitive to the attribute data that is missing. For example, a bronze subscription user or a web browsing service may not be sensitive to small variations in QoS related attribute values. This is primarily because of the low weights assigned to the QoS related attributes in the MADM decision process for such users/services. If the sensitivity is low, a standard MADM algorithm can be applied even with imprecise attribute information. The association of services and subscription types to different levels of sensitivities has been considered in Chapter 3. However if the sensitivity is high or it cannot be determined because of missing service/subscription related information then data estimation should be done. Depending upon the estimation methods used and the past experiences with the estimation process, the forecasted data can be expressed as a scalar value or as a fuzzy set. For fuzzy data values, the fuzzy implementation of a MADM algorithm is then applied to get a fuzzy ranking of the networks. This is followed by the step where the fuzzy rankings are defuzzified. For the case of non-fuzzy estimated values, a standard MADM algorithm is applied. In the final step, any additional information to judge the Confidence Level in the network ranking obtained is taken into consideration. In another approach described later in the chapter, which has some advantages to the approach in Figure 5.3, the step involving Confidence Level has been moved to after forecasting of the missing data.

5.3.1.1 Data Estimation

A uniform set of attributes has to be available as inputs to the decision algorithm that would compare different access networks on their basis. Data prediction can be used to see what will happen to attribute values in the future based on the past history of the data. For example, based on the time of day and usage pattern, many of the QoS attributes such as the utilization of a hotspot can be predicted to a certain degree. Typically the observed values are represented in the form of time series which are collections of observations ordered in time. Deterministic data prediction can be performed using the following methods.

Seasonal trends and averaged values

Smoothing techniques can be used to identify the underlying trends in the observed values of attributes when some of the data to be used in the decision making process are missing. The
smoothing of time series removes noise related irregularities and enhances the informational part of the observed data. Several variations of smoothing techniques exist; e.g., Simple Moving Averages (SMA) is useful for the type of data that exhibits static values for the mean and variance. SMA of order \( n \) at time \( t+1 \) can be described by the following formula,

\[
SMA_{t+1} = \frac{(V_t + V_{t+1} + \ldots + V_{t-n+1})}{n}
\]

Weighted Moving Averages (WMA) involves assigning different weights to historical data before taking the moving average. It can be represented as follows,

\[
WMA_{t+1} = \frac{(w_i \cdot V_t + w_{i+1} \cdot V_{t+1} + \ldots + w_{n+1} \cdot V_{t-n+1})}{n}
\]

where

\[
\sum_{i=t-n+1}^{i=t} w_i = 1
\]

Other types of moving averages include various forms of Exponential Smoothing (ES) techniques described in [12]. The use of a particular smoothing technique would depend upon its ability to accurately predict the attribute values. The accuracy of prediction can be measured by running the algorithm on prior collected data and comparing the actual and the predicted values.

**Regression**

Regression [12] analyzes the relationship among variables to estimate the value of one variable from known values of other variables related to it. In the case of network selection, an attribute with known value can be used in calculating the value of another input attribute because of a strong correlation between the two attributes. In regression analysis trends of variables under consideration are analyzed (such as linear) and the variables that are seen to have dependencies on each other are correlated and then modeled using polynomials. The variables that can be easily measured with least error are also identified. A regression using only one predictor is called a simple regression. For example if network utilization is easy to measure and report but packet jitter, packet loss, or latency are not readily available, then these QoS related parameters can possibly be estimated by the network utilization because of a correlation between them under normal network conditions.
The actual process of finding the relationship can involve sample data collection, drawing scatter plots to understand the relationship amongst the variables, and using computer packages or modeling techniques (such as linear prediction using least square method) to figure out the relationship. Typically the established relationship only holds for a limited range of variable values and only provides an estimate or average value.

As an example, an access network under consideration maybe able to provide its network utilization more readily, reliably and on a continuous basis than other QoS related parameters such as packet loss, delay and jitter, which can require much more active monitoring of the network. Under normal operating conditions the network utilization is correlated with QoS aspects in a packet-switched network, using regression analysis as described in this section. Therefore the utilization (L) that is much easier to monitor and to report on, can be related over a range of values to the packet Delay (D), Jitter (J) and Loss (L). Similar relationships between network utilization and QoS parameters may be provided by the operator of the network to its partners as part of the roaming agreements or SLAs.

**Fuzzy Estimated values**

Because of the forecasting techniques used and other factors such as the prior experience with the use of forecasted data, non crisp values may be obtained which can then be represented as Fuzzy Numbers. A Fuzzy Number [4][5] forms a fuzzy set that can have different membership grades as shown in Figure 5.4. A very common type of fuzzy set membership results in a Triangular Fuzzy Number (TFN). With the type of problems that can arise in data acquisition such as measuring inaccuracies or lack of updated real time attribute data as described earlier, a triangular shaped fuzzy number will be a good fit. For example when regression analysis is used for data prediction, the value of the predicted parameter can be described by a triangular fuzzy number. Figure 5.4 also provides the mathematical representations of the membership function for a TFN and some simple operations like addition and multiplication between two TFNs.
Simple Fuzzy Number Sets

\[ \mu_a(x) = \begin{cases} \frac{x - a}{b - a}, & x \in [a, b] \\ \frac{x - c}{b - c}, & x \in [b, c] \\ 0, & \text{otherwise} \end{cases} \]

Membership function for triangular fuzzy number

**Figure 5.4 Fuzzy Numbers**

Use of tools such as scatter plots shows that in general the correlation defined by regression between the variables is never exact. For example, in the case of the regression scenario described earlier, at lower utilization there will not be any congestion conditions in the AN and therefore it can provide a high level of QoS to all the services or all IP flows. In this case the regression equations would provide an accurate relationship. At a higher utilization the AN could distinguish between preferred and non preferred services or IP flows in the packet scheduling process. As a result, the aggregate observed data for all services/IP Flows would show a spread/scatter at higher utilizations. This behavior can be represented by mapping the attribute values of D, J and L to a range of U values. For this example we consider that they map to 0.8-1.1U with 1.0U representing the default or modal value.

In fuzzy numbers terms, this can then be represented by triangular fuzzy numbers shown graphically in Figures 5.4. It can be seen that uncertainty in the values for D, J and L increase with an increase in the network utilization (U). Fuzzy number sets can also be obtained by direct mapping from a lookup table that maps a known value of U to D, J, L values observed earlier for the same value of U in that network (as described by seasonal tends in the previous section).
5.3.1.2 Compensating MADM Algorithm

MADM algorithms include a variety of deterministic mechanisms [1][2] that have been used in decision making for a wide variety of engineering problems. Fuzzy logic can be applied in conjunction with a MADM algorithm. Triantaphyllou in [6] and [7] has described the use of fuzzy information with some of the commonly used MADM algorithms.

Table 5-1 Attributes and their weight assignments for streaming media and web browsing services while using GRA

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Streaming Media</th>
<th>Web Browsing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost per Byte (CB)</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>i.e. Data transport cost on a particular access system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Bandwidth (TB)</td>
<td>0.15</td>
<td>0.05</td>
</tr>
<tr>
<td>i.e. Overall bandwidth of the wireless access link</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allowed Bandwidth (AB)</td>
<td>0.2</td>
<td>0.15</td>
</tr>
<tr>
<td>i.e. Bandwidth per user allowed by the access system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilization (U)</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>i.e. Current utilization of the wireless link</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Packet delay (D)</td>
<td>0.1</td>
<td>0.05</td>
</tr>
<tr>
<td>i.e. Average packet delay within the access system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Packet Jitter (J)</td>
<td>0.1</td>
<td>0.05</td>
</tr>
<tr>
<td>i.e. Average packet delay variations within the access system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Packet Loss (L)</td>
<td>0.05</td>
<td>0.1</td>
</tr>
<tr>
<td>i.e. Average packet loss rate within the access system</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5.5 Example graphs of packet Delay (D), Jitter (J) and Latency (L) values using Network Utilization (U). It assumes correlation represented by $D = 0.225U^2 + 10$, $J = 0.05U^2$, $L = 0.0019U^3$ with 0.8U - 1.1U in the network utilization range of 5% and 70%.

Depending upon the type of ambiguity in the decision process, fuzzy logic can be applied in more than one ways with MADM. For example, the uncertainty in the decision process can be in the
attribute values or it can be in the importance associated with them. Here, it is assumed that the importance or weight assigned to an attribute is based on the service requested or the user’s QoS subscription and therefore known based on the information provided by the user. However, because of the reasons described earlier in the chapter, some of the input attribute values are either unavailable or are imprecise.

Here GRA [13][14] is used as a MADM algorithm for fuzzy implementation described in this chapter. GRA [14] is based on the concept of GRS, where GRS \((X,Y)\) describes a relationship \(Y\) between reference data values \(X_0\) and a sequence of data values \(X\). So if \(y \in Y\), \(x_i \in X\), \(x_0 \in X_0\) such that \(x_0 = x_0(n)\) and \(x_i = x_i(1)\), \(x_i(n)\) then \(y(x_0(k),x_i(k))\) would represent a GRS at point \(k\) provided the axioms documented in [14] are satisfied. GRA calculates a GRC and uses it as a measure of the closeness of the result to the reference values. The GRC values can thus be used for ranking the candidate networks.

**Table 5-2 Attribute values for the five networks under consideration**

<table>
<thead>
<tr>
<th>Ntwk#1</th>
<th>CB (%)</th>
<th>TB (mbps)</th>
<th>AB (mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>2</td>
<td>0.2</td>
<td>10</td>
<td>400</td>
<td>50</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ntwk#2</th>
<th>CB (%)</th>
<th>TB (mbps)</th>
<th>AB (mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>11</td>
<td>1</td>
<td>20</td>
<td>200</td>
<td>25</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ntwk#3</th>
<th>CB (%)</th>
<th>TB (mbps)</th>
<th>AB (mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>54</td>
<td>2</td>
<td>20</td>
<td>100</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ntwk#4</th>
<th>CB (%)</th>
<th>TB (mbps)</th>
<th>AB (mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>100</td>
<td>5</td>
<td>40</td>
<td>150</td>
<td>30</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ntwk#5</th>
<th>CB (%)</th>
<th>TB (mbps)</th>
<th>AB (mbps)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>100</td>
<td>5</td>
<td>20</td>
<td>100*</td>
<td>20*</td>
<td>15*</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>122'</td>
<td>25'</td>
<td>25</td>
<td>25</td>
<td>24'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>151'</td>
<td>31'</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>165'</td>
<td>34'</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*- last reliable attribute values

`- Forecasted attribute value represented as a Triangular Fuzzy Number

In applying the fuzzy implementation of GRA to the scenarios under consideration the attribute values to be used as inputs to the algorithm will be a mixture of crisp and fuzzy values. This is because not all of the attribute values are unknown or have imprecise values. For networks with fuzzy input attributes represented by TFNs, it would result in three GRCs to be evaluated using the fuzzy arithmetic for TFN as shown in Figure 5.4. This assumes that because of correlation
amongst missing input attribute values, multiple fuzzy inputs described by TFNs will only result in three possible input combinations. A defuzzification process is then used to get a crisp value of the GRC.

In order to better understand the network selection decision process described in Figure 5.3 we consider a simple scenario where a network selection decision has to be made with five network options. These networks are represented by a set of QoS related attributes as listed in Table 5-1. The selection of these attributes for use in network selection while using MADM is described in Chapters 3 and 4 [17][18][19]. Each attribute has a relative importance represented by a weight assigned to it during the decision process. The relative importance of an attribute during the decision process can be based on the service being requested or the QoS profile of the user. The weights shown in Table 5-1 represent the importance assigned to the attributes for a streaming service and web browsing. We have selected streaming service as a representative of a service scenario that is sensitive to dynamic QoS attributes and web browsing as a representative service scenario that is not sensitive to dynamic QoS attributes.

![Figure 5.6 Results for streaming media services using GRA algorithm with fuzzy input and with imprecise input information](image-url)
Table 5-2 lists the values for these attributes for Ntwk#1 through Ntwk#4 at the instance of decision making. QoS related data (i.e., Delay, Jitter and Packet Loss) for Ntwk#5, one of the networks under consideration, is not available at the time of decision making. However the last known reliable values for these attributes for Ntwk #5 is available and are listed in Table 5-2. We assume that based on prior observation of the network, it has been possible to correlate network utilization with QoS parameters. The correlation used in this example is defined in Figure 5.5. It results in TFNs for Delay, Jitter and Latency values for Ntwk#5 as shown in Table 5-2.

Using the fuzzy version of GRA described in the previous section for the streaming media service, the results as shown in Figure 5.6 are obtained. It can be seen that for Ntwk#5, three GRC values are obtained which constitute a GRC represented by a TFN. With Ntwk#5 represented by a fuzzy GRC, the exact ranking for the networks is not obvious and therefore a defuzzification step is required. It should be noted that while it is observed in Figure 5.6 that changing the attribute values for one network, in this case Ntwk#5, does not impact the GRC values of the other networks, this is not always the case. In most of the cases, the way the GRC is calculated (as shown by its mathematical representation earlier), some of the other network alternatives may also result in different GRC values. when the attribute values for one network is changed. In that case the defuzzification part can become more complicated as we will have to apply defuzzification techniques to other network alternatives as well in order to obtain crisp rankings. This would become clearer in the example presented in the next section.

5.3.1.3 Defuzzification

Defuzzification is the process of converting a fuzzy number or set to a crisp value. Several mechanisms exist for comparing fuzzy numbers with some of them being fairly complex and computationally intensive. A good mechanism can be selected by considering its level of complexity, its accuracy and its application to fuzzy numbers of a particular shape. [15] [16] documents a range of defuzzification techniques. Center of gravity (CoG) is a very popular technique that is computationally among the simplest and particularly well suited for defuzzification of TFNs. It calculates the centroid of the fuzzy value and can be approximated in the discrete domain as follows,

\[ CoG \approx \frac{\sum_{j=1}^{r} \mu_j * F_j}{\sum_{j=1}^{r} \mu_j} \]
where $\mu_j$ and $F_j$ represent the fuzzy set membership index and the corresponding fuzzy number, respectively, in the discrete domain. A simple Matlab implementation of the algorithm for calculating the CoG of a TFN for GRC is shown below:

```matlab
for F=a:increment:b
    U1 = F/(b-a)-a/(b-a);
    SumofU1 = U1 + SumofU1;
    SumofUF1 = F*U1 + SumofUF1;
end
for F=b:increment:c
    U2 = F/(b-c)-c/(b-c);
    SumofU2 = U2 + SumofU2;
    SumofUF2 = F*U2 + SumofUF2;
end

CoG = (SumofUF1 + SumofUF2)/(SumofU1 + SumofU2)
```

Figure 5.7 graphically shows application of CoG for defuzzification of the GRC value for Ntwk#5. Comparing the CoG value for Ntwk#5 from Figure 5.7 with the results for the remaining networks as shown in Figure 5.6 it can be seen that the CoG based GRC value for Ntwk#5 is lower than the GRC value for Ntwk#4. Hence Ntwk#4 should be selected.

In the absence of reliable data for some of the networks, the alternatives to not using data forecasting and fuzzy techniques as described above would be to:

1. Apply a standard MADM technique for all the networks without fuzzification and use instead non-fuzzy unreliable data or their approximations as inputs.

2. Remove the networks with unreliable or missing data altogether from the comparison and then perform MADM on the remaining networks to obtain their rankings.
If the first approach is used, depending upon the sensitivity of the selected service and/or the user subscription type to the missing information, there is a possibility of getting an incorrect ranking of the networks because of unreliable inputs to the algorithm. For streaming media service this is shown in Figure 5.6 where the last reliable values of missing parameters were used. However the values were outdated and since then network utilization had gone up. So because of a strong correlation between network utilization and the missing QoS attributes, the values of missing attributes should have been predicted since streaming media service is sensitive to QoS attribute values. The use of outdated values in this case leads to the incorrect selection of Ntwk#5 whereas Ntwk#4 would have been a better choice as shown by the use of the fuzzy GRA implementation also in Figure 5.6.

![Triangular Fuzzy Set](image)

**Figure 5.7 Use of CoG for defuzzification of a TFN representing the GRC in a fuzzy implementation of GRA for streaming media services**

If the second approach is used, it is possible that a top ranking candidate network is not selected as it is dropped before the comparison process. For the same example described above, the results of applying GRA after removing Ntwk#5 from the candidate list are shown in Figure 5.8. While in this case Ntwk#4 is correctly selected, if in the given example, the network utilization had gone down instead then it is entirely possible that Ntwk#5 would have been a better choice but it would have been eliminated as a candidate by this strategy.
Figure 5.8 Results of GRA algorithm for streaming media service after removing Ntwk#5 with missing information

Figure 5.9 Results for web browsing service using GRA algorithm with fuzzy input information and imprecise input information
Now we consider the case of web browsing service. In the case of this service, based on the weights assigned to the attributes the sensitivity of the service is known to be low to dynamic QoS attributes. In order to confirm our assertion that for this type of service the use of imprecise information is acceptable, we apply both fuzzy (with parameter estimation) and crisp versions of the GRA algorithm. The results of the fuzzy GRA algorithm with forecasted data documented in Table 5-2 are shown in Figure 5.9. The results of defuzzification are shown in Figure 5.10. The results of application of crisp GRA while using imprecise information for the missing attributes are shown also in Figure 5.9. Comparison of the two set of results shows that Ntwk#4 gets selected in both cases by a wide margin and the use of imprecise information does not have much impact on the results in this case.

Figure 5.10 Use of CoG for defuzzification of a TFN representing the GRC in a fuzzy implementation of GRA for web browsing services

123
An appropriate use of parameter estimation and the fuzzy techniques as described in Figure 5.3 hence provides a balanced approach by not eliminating the networks with missing, old or unreliable data while keeping in view the fact that the information about them is fuzzy and/or not entirely accurate. In the following section the proposed mechanism has been further enhanced by developing criteria for judging the reliability of the results.

5.3.1.4 Confidence Level

Although the methods described above can provide improved ranking for the networks under consideration, other factors should also be taken into consideration before making use of the rankings obtained in this manner. Here we introduce the new concept of "Confidence Level" that can allow the decision maker to evaluate if the network rankings are reliable enough to be used for network selection. The factors used in calculating the Confidence Level (CL) include:

- **Sensitivity of the service or user subscription type to unreliable attribute values (CL(s)):** As described earlier, in the application of a MADM process, attribute weights are assigned based on the service or user type; e.g., a VoIP service or a Gold user may be assigned higher weights for dynamic QoS related attributes because of the nature of the requested service or the user subscription. Based on the weights assigned to the attribute values, network ranking can become more vulnerable to errors in parameter estimations. In other words, a decision where dynamic QoS attributes have higher assigned weights will have a higher sensitivity to errors in the parameter values. Figure 5.11 shows the typical relationships between CL and some of the common service types while Figure 5.12 shows the relationships between CL and the reliability of data values for the same services. For example a VoIP service when compared to web browsing is more vulnerable to incorrect ranking because of imprecise values of the QoS related attributes. This is because of the relatively higher weights assigned to the dynamic QoS attributes such as packet delay, jitter, and loss in decision making for QoS sensitive service types.
Figure 5.11 Trend graph for relationships between Confidence Level and common service types when a QoS related parameter with dynamic characteristics has unreliable value and therefore is being predicted.

Figure 5.12 Trend graph for relationship between Confidence Level and reliability of estimated data for some of the common service types.

- **Time since the last data update (CL(t))**: There will be a level of uncertainty in the data values based on how long data have been unavailable and how much they can change over time based on, e.g., seasonal charts. Figure 5.13 shows a typical relationship between CL and the...
time since the last attribute value was received indicating the older the last good value, the lesser the confidence in its estimated value.

- The degree of correlation with attribute whose value is known and is used in the regression analysis (CL(c)): This indicates the level of confidence in the estimated value. A typical relationship between the CL and the degree of correlation obtained from regression is shown in Figure 5.14. A strong correlation would indicate a high confidence level.

![Figure 5.13 Trend graph for relationship between Confidence Level and time since last attribute value was received](image1)

**Figure 5.13 Trend graph for relationship between Confidence Level and time since last attribute value was received**

![Figure 5.14 Trend graph for relationship between Confidence Level and degree of correlation between known and unknown attributes](image2)

**Figure 5.14 Trend graph for relationship between Confidence Level and degree of correlation between known and unknown attributes**
We define a minimum threshold (CL\textsubscript{threshold}) of CL for an acceptable level of the ranking results. In other words, for the ranking to be considered trustworthy, the following should be true.

\[ CL > CL_{\text{threshold}} \]

The value of CL\textsubscript{threshold} can be provisioned into the process by the decision maker based on the decision policies. In the case of CL being lower than the threshold, the decision maker will remove the network with imprecise information from the list of alternatives being considered. However, it is known [7] that addition or removal of an alternative during application of a MADM algorithm can impact the ranking for the rest of the alternatives non-uniformly. It may therefore be necessary to run the MADM algorithm again in this situation. An alternative approach is to calculate CL before using the MADM algorithm. If CL is found to be below the threshold for all the alternatives with imprecise information, then they are removed from the comparison and a standard MADM algorithm is applied to the rest. Otherwise only those alternatives with CL lower than the threshold are removed and the fuzzy version of the MADM algorithm is applied to the rest. This approach is shown in Figure 5.15.

![Figure 5.15 An alternative approach of using Confidence Level in the decision process](image-url)
Using the alternative approach and the factors described above, the decision maker can formulate an overall CL to be used in the decision process. A very simple example of such a term can be

\[ CL = CL_s + CL_t + CL_c \]

where \( 0 \leq CL \leq 1 \).

The decision maker can also provide, based on his understanding of how various factors should influence CL, some distribution of weights across CL\(s\), CL\(t\) and CL\(c\). Here we use an equal distribution of weights between the three so that \( 0 \leq CL_s \leq 0.33 \), \( 0 \leq CL_t \leq 0.33 \) and \( 0 \leq CL_c \leq 0.33 \).

Using the trend graphs shown in Figure 5.13 and Figure 5.14, assuming a linear relationship between the variables and utilizing the normalized values, we come up with the following two equations,

\[ CL_t = -0.33t + 0.33 \]

where \( t \) is the normalized time, i.e., \( 0 \leq t \leq 1 \) with \( t = 1 \) representing a finite elapsed time since the last reliable value after which data values can no longer be considered useful.

\[ CL_c = 0.33c \]

where "c" is a normalized correlation factor, i.e., \( 0 \leq c \leq 1 \) determined by the decision maker based on past data collection.

Similarly we can relate CL\(s\) to the normalized value of sensitivity factor determined by the decision maker as follows

\[ CL_s = 0.33s \]

where \( 0 \leq s \leq 1 \).

Using these values for CL\(s\), CL\(t\) and CL\(c\) we come up with a simple equation for CL as follows

\[ CL = \frac{(s + c - t + 1)}{3} \]

where s, c, and t are all normalized values.
Figure 5.16 Graphical representation of how network congestion during service delivery time can be related to parameters such as the total bandwidth, bandwidth allowed per user, current network utilization and rate of change of network utilization.

In addition, deduction from the information available about other attributes can help the final network selection process. One such example is shown in Figure 5.16 where the possibility of network congestion during service delivery can be predicted based on service related information, the available bandwidth and total bandwidth for the network under consideration, the current network utilization and its past rate of change. Such information can then be used in the final selection process. Another application of the type of prediction shown in Figure 5.16 can be when the duration of the service to be used is known. In that scenario, it can be possible to use the predicted parameter values into the future and apply network selection in future at more than one time instances to see if the same network would be selected for the duration of the service. This type of application would be especially useful if the networks under consideration do not have good admission control and QoS guaranteeing mechanisms in place and therefore service experience for already admitted sessions can suffer from increased network load.

Since forecasting of data and CLs of the estimated values can be done independent of the use of fuzzy MADM algorithm, it is also possible for the decision maker to update the CLs of the estimated values on a regular basis and make them available to the decision process when needed. A lookup table for each of the factors impacting the CL of each attribute can be created to map the
normalized values of these factors to a CL for easy calculation. A variety of decision types, from a rough estimate to very accurate, can be supported depending upon the resolution or frequency of updates of the data being forecasted. In order to better understand the notion of *Confidence Level* we consider the use of lookup tables shown in Table 5-3 with the example described previously in the chapter. The values used in the table are for illustration and will be decided in practice by the decision maker based on the decision policies and the past experience with data prediction. For example, the table entries for CL(s) indicate the fact that an unreliable Packet Jitter value will result in a lower CL for VoIP type service when compared with Web Browsing. Also in Table 5-3, it is assumed that CL(t) and CL(c) values for Packet Delay, Jitter and Loss will be the same as these parameters are very closely related.

Table 5-3 Lookup tables used by the decision maker for Confidence Level calculations

<table>
<thead>
<tr>
<th>Service Type</th>
<th>CL(s) for Packet Delay, Jitter and Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>VoIP</td>
<td>0.5</td>
</tr>
<tr>
<td>Streaming Media</td>
<td>0.75</td>
</tr>
<tr>
<td>Web Browsing</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time since last Reading (sec.)</th>
<th>CL(t) for PacketDelay, Jitter and Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>600</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute</th>
<th>CL(c) with Network Utilization (U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet Delay (D)</td>
<td>0.9</td>
</tr>
<tr>
<td>Packet Jitter (J)</td>
<td>0.9</td>
</tr>
<tr>
<td>Packet Loss (U)</td>
<td>0.9</td>
</tr>
</tbody>
</table>

We set the threshold for CL at 0.5. We assume that at the time of decision, the attribute values had not been updated for the past 5 minutes. Table 5-3 shows that in this case it has been
determined that the data values are not useful beyond 600 seconds or 10 minutes. This would therefore give us a normalized value of $t=0.5$. Since the service type in the example was streaming media, the normalized value of $s$ is 0.75. Also per Table 5-3, the normalized correlation factor $c$ for missing data is 0.9. Using these values in the prior equation we come up with a value for $CL$ of 0.717 which is higher than the threshold and therefore allows for inclusion of the access type in network selection process. This type of simple calculations can help improve the reliability of results by incorporating the decision maker's judgment under less certain conditions.

5.3.2 Scenario 2 - Network Types with Non Crisp Attributes

Some of the candidate networks may have a range of QoS values instead of one crisp value. This can be for different reasons:

- A network may support multiple QoS classes or SLAs with different cost structures associated with them as described in Chapter 3. These QoS classes / SLAs can be treated as separate alternatives while using a compensating MADM algorithm. The less expensive QoS classes / SLAs may have more variability on the QoS attributes and such variability can be represented by a nominal value for each attribute with a possible spread of values around it as represented by a TFN described earlier.

- The access technologies being used by some candidate networks may be inherently incapable of providing strict QoS guarantees. This can also be based on the decision maker's prior experience with a operator network or access technology. The variability of this type can be represented by a fuzzy number.

If the service or subscription information is considered sensitive to the non crisp attribute values then network selection in such situation can make use of a fuzzy MADM algorithm as described in the previous section. Steps related to data harmonization, fuzzification and estimation of $CL$ would not be required. The fuzzified input data is assumed to be available in this case (e.g., pre-provisioned) and after application of the fuzzy MADM the results are defuzzified to get crisp rankings.
In order to better understand this type of usage of MADM algorithms we consider a network selection scenario where multiple QoS/SLAs with a network operator constitute the alternatives, similar to the scenario described in Chapter 3. Further we assume that the user plans to use streaming media service with the attribute weights as shown in Table 5-1. The attribute values for the QoS/SLAs classes are shown in Table 5-4. One of the less expensive alternatives, Ntwk#3 as shown in Table 5-4 does not provide crisp QoS attributes which indicates that there can be significant variation in the values during a ongoing session depending upon congestion levels.

Applying the fuzzy GRA algorithm using the process described in the earlier section we get the results shown in Figure 5.17. As was discussed earlier, a change in attribute values for one of the network can change the values of the GRCs for other networks as well. This is shown in the results in Figure 5.17. In order to get crisp values in this case, defuzzification is applied to the GRC values of all the network alternatives. The results of defuzzification process are shown in Figure 5.18. The crisp values obtained from Figure 5.18 indicate that Ntwk#3 has the maximum GRC value and is the preferred network in this case.

<table>
<thead>
<tr>
<th>QoS Class/SLA #</th>
<th>CB (Mbps)</th>
<th>TB (Mbps)</th>
<th>AB (%)</th>
<th>U (%)</th>
<th>D (msecs)</th>
<th>J (msecs)</th>
<th>L (per 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QoS Class#1/SLA#1</td>
<td>10</td>
<td>100</td>
<td>0.5</td>
<td>30</td>
<td>400</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>QoS Class#2/SLA#2</td>
<td>20</td>
<td>100</td>
<td>1</td>
<td>20</td>
<td>200</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>QoS Class#3/SLA#3</td>
<td>30</td>
<td>100</td>
<td>1</td>
<td>20</td>
<td>100</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>QoS Class#4/SLA#4</td>
<td>50</td>
<td>100</td>
<td>2</td>
<td>40</td>
<td>50</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>QoS Class#5/SLA#5</td>
<td>60</td>
<td>100</td>
<td>5</td>
<td>60</td>
<td>40</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Figure 5.17 Results of fuzzy GRA algorithm
Figure 5.18 Use of CoG for defuzzification of TFN representing GRCs in the fuzzy implementation of GRA
5.4 Conclusion

Selection of an optimal service delivery network is an important problem to be solved in an all-IP heterogeneous wireless networking environment spanning multiple operator domains. The problem requires special consideration when attribute related data is unreliable or unavailable. Prior work in this area has limited applicability as it has not looked at the problem in its entirety. For example prior studies did not take into consideration important aspects such as the sensitivity of the network selection to imprecise information, data prediction for unavailable data, use of fuzzy number types to represent predicted attributes and the decision strategy to determine when to apply fuzzy MADM algorithms. In the absence of a well defined strategy to handle such scenarios, suboptimal networks could be selected. In this chapter we have provided a comprehensive network selection mechanism in the presence of imprecise information. We have considered scenarios where fuzzy techniques along with data prediction can be applied in the network selection decision process while using a fuzzy implementation of the Grey Relational Analysis MADM algorithm. We have developed additional decision support techniques that can be applied under such conditions. These techniques have been used to obtain the Confidence Levels in the network rankings obtained via the use of fuzzy logic and data prediction with a MADM algorithm. An alternative approach for removing low confidence alternatives earlier on from the decision process has also been described. The techniques described in the chapter can be tuned to address different decision types and resolutions of the data. Through examples we have shown that the techniques proposed in this paper can help improve the reliability of the results and allow for improved network selection under uncertain conditions.
5.5 References


6 Conclusions

Inter-working of heterogeneous wireless networks has emerged as a commercially viable solution to the problem of efficient and ubiquitous delivery of services to nomadic and mobile users. Network selection is a critical problem to be addressed in interworked heterogeneous all-IP wireless systems. It has both revenue and service impacts. In our research we have provided solutions to address this problem. The fact that several industry forums are actively looking to solve this problem with multifaceted requirements indicates its importance for the industry. The issue is of special significance as it impacts both the operator's revenue and customer's service experience. The decision algorithms used in the networks are unlikely to be standardized but used by the operators and equipment manufacturers as means of differentiating their services or products. However many of the architectural ideas presented in the earlier chapters of the thesis are a topic of discussion in industry forums and standards bodies. IEEE 802.21 continues to work on Media Independent Handover (MIH) solutions, part of which is to solve the problem of network selection across heterogeneous network technologies during inter-technology handoffs. More standardization efforts would be needed to enable 802.21 type mechanisms in prevalent wireless wide area access technologies. The IEEE 802.11 and IEEE 802.16 groups are currently working to enable MIH based on IEEE 802.21 work and similar efforts maybe needed in other forums such as 3GPP. 3GPP independently is also looking into the problem of network discovery and selection in System Architecture Evolution as part of supporting seamless mobility across multiple access technologies. In fact as a validity of the architectural approach proposed here, various industry forums have realized the interference issues with two radio solutions during network selection process for seamless inter-technology handoffs and are now considering single radio solutions where one access technology network can provide information about networks with other access technologies available in the vicinity.

We conclude this Chapter with a summary of the contributions of the thesis and future work that is possible in the area.

6.1 Summary of Contributions

Architectural Framework

Currently there is no good deployable solution in the industry to deal with the problem of network selection in a heterogeneous all-IP wireless networking environment. Our research and
in turn contribution to the work\textsuperscript{1} [1] being done at IETF has provided a better understanding of the problem space and also established design constraints/recommendations for an acceptable solution. Using these guidelines we have proposed an architectural framework for network selection in a heterogeneous wireless networking environment in [2] and [3] and described it in Chapter 2. In our proposed framework, the network assists the terminal in the decision process. The proposed solution scales well, is flexible and also supports roaming. It. minimizes the usage of the wireless links and would work with currently deployed infrastructures. The architectural solution we have proposed in the thesis leverages the concept of Information Services defined in IEEE 802.21. Using IS at layer 3 via a client/server model is a notion that we have contributed to in IEEE 802.21\textsuperscript{2}[4] as well. We have also described how information exchange using layer 2 can occur in our architecture in the case of WLANs. The IEEE 802.11u based layer 2 network selection system referred in the example in Chapter 2 is a proposal to which we have also contributed\textsuperscript{3} [5] in IEEE 802.11u. Finally we have integrated our architecture with the current

\textsuperscript{1} The author of this thesis was one of several co-authors to the referred industry standards submission and therefore does not claim authorship of the entire referred document or the ownership of all the ideas presented in it. See reference for the complete list of authors.

\textsuperscript{2} The author of this thesis was one of several co-authors to the referred industry standards submission and therefore does not claim authorship of the entire referred document or the ownership of all the ideas presented in it. See reference for the complete list of authors.

\textsuperscript{3} The author of this thesis was one of several co-authors to the referred industry standards submission and therefore does not claim authorship of the entire referred document or the ownership of all the ideas presented in it. See reference for the complete list of authors.
cellular infrastructure by proposing to enhance the function of some of the existing functional
nodes (PDF and AAA) in the cellular system and adding new logical information nodes (SAN
and DCN).

**Comprehensive MADM based Network Selection Decision Process**

Prior work on the network selection decision process has not considered some of the key
decision attributes, e.g., multiple authentication credentials, multiple authentication mechanisms,
user's mobility profile, and roaming relationships between the user's HN and visited networks.
This can be attributed to lack of understanding of the problem space, unavailability of an
architectural framework to provide a solution in, and the inability to use one decision algorithm
for all attributes that needed different type of decision making. Within the framework of our
proposed architectural framework, we have developed a two step decision process for a network-assisted
network selection mechanism that combines non-compensatory and compensatory
MADM algorithms [6].

**Improvements to MADM algorithms for application to network selection when complete
information about the factors impacting network selection is available**

During the course of our research we have proposed improvements to existing MADM
algorithms or improvement towards their application to this problem space. We have identified
causes of ranking abnormalities in TOPSIS and proposed an improvement to the standard
TOPSIS algorithm when it is applied to network selection by adopting an iterative approach [7].
The need for support of different optimization objectives for the decision maker (e.g., the
network operator) and hence the need to support non-monotonic utility for some of the attributes
used in decision making has been demonstrated [8]. We have shown that many of the commonly
used MADM algorithms such as SAW, WPM, TOPSIS, and ELECTRE in their standard forms
are not well suited to network selection because of their assumptions about monotonically
increasing or decreasing utility of each attribute. It has been shown that GRA is better suited for
achieving this type of optimization objectives with the selection of delivery network impacted by
how the algorithm is used to achieve the optimization objectives. A two step process has been
proposed and evaluated to show how reference attribute values and attribute weights impact the
selection process. The adjustment of these parameters for different service types has been
discussed. ELECTRE, a well known MADM algorithm, has been modified [9] to allow its usage
for network selection with attributes exhibiting a non-monotonic utility. This modification makes
the algorithm more adept to application in network selection as it expands its applicability to a wider range of optimization objectives. To the best of our knowledge the enhancements to the MADM algorithms themselves or to their application in network selection listed above are the first. In addition the improvements to the algorithms are not specific to the problem of network selection but can be applied to other problem spaces with similar requirements.

Application of data prediction and fuzzy techniques with MADM based network selection for scenarios when some of the information to be used in network selection is missing

Prior studies did not develop a decision strategy about when to apply standard MADM algorithms and when to apply fuzzy MADM algorithms in the presence of imprecise input information. We have proposed such a decision process for network selection with imprecise input information [10], which considers when fuzzy techniques along with data prediction can be useful in network selection decisions. We have proposed a new function called the *Confidence Level* for network selection decision support. Together with the sensitivity of the missing information, the two functions help determine if it would be useful to estimate the data, if the estimated data should be used in the decision process, and if the results thus obtained could be trusted. To the best of our knowledge this is the first such proposal for solving the problem of network selection with imprecise input information.

The decision entity applying the MADM algorithm has been assumed to be in the network in our proposed architectural framework. However the application of the algorithms presented in the thesis to network selection is irrespective of the architectural framework. The algorithms described here would work equally well if the decision making entity resides in the terminal.

### 6.2 Future Work

In the coming years ideas presented in this thesis will be validated or challenged in the market place. Along with technical viability of the solutions presented here it will be the new and existing business models used by the network operators that will determine the feasibility of solutions presented in the thesis. For example the idea of sharing network related information across partner networks although technically feasible will eventually depend on how willing the operators are in sharing their network related data. More work in the areas of Dynamic SLAs and Game Theory can help extend the application of these techniques to new business models. From the algorithmic perspective, the focus of the thesis has been on the use of deterministic decision making techniques. This has allowed for ease of implementation that is needed for real time decision
making and for the ease of understanding of the underlying decision making philosophy by the
decision maker, e.g., the network operator. The possibility of using non-deterministic algorithms
(e.g., use of Q-learning) can also be explored in future research while keeping in mind the
requirements for ease of implementation and ease of understanding of the decision making
philosophy. In our research we have assumed a central decision making entity located in the
network. It is possible that the decisions are made in steps and at more than one locations. Thus a
possible future research area can be in a multistage decision making approach that involves
decisions by more one entities distributed in their locations.
6.3 References


