A Service Provisioning Platform for Beyond 3G - The Telecommunication Service Exchange

Donna Griffin

Department of Electronic Engineering, Cork Institute of Technology, Cork, Ireland.

Follow this and additional works at: https://sword.cit.ie/allthe

Part of the Electrical and Electronics Commons, and the Systems and Communications Commons

Recommended Citation
Available at: https://sword.cit.ie/allthe/382

This Doctoral Thesis is brought to you for free and open access by the Dissertations and Theses at SWORD - South West Open Research Deposit. It has been accepted for inclusion in Theses by an authorized administrator of SWORD - South West Open Research Deposit. For more information, please contact sword@cit.ie.
A Service Provisioning Platform for Beyond 3G -
The Telecommunication Service Exchange

Donna Griffin
A Service Provisioning Platform for Beyond 3G – The Telecommunication Service Exchange

Donna Griffin M.Eng., B.Sc.
Department of Electronic Engineering
Cork Institute of Technology

Supervisor
Dr. Dirk Pesch

Submitted in partial fulfillment for the Degree of
Doctor of Philosophy

Submitted to Cork Institute of Technology, September 2007
Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material to a substantial extent has been accepted for the award of any other degree or diploma by a university of higher learning, except where due acknowledge has been made in the text.

Signature of Author: ........................................

Certified by: ........................................

Date: ........................................ 12/10/2007
From the inception of mobile communications one of the most important factors contributing to its success has been a good understanding of user requirements. In the first generation, mobile networks offered subscribers the freedom to communicate regardless of their locations and the joy of instant access to voice telephony services. This provided a competitive edge over wire line communication networks and resulted in large revenue streams. However over the past decade, mobile network operators have been faced with a number of factors that have incurred significant costs and resulting reduced revenue. These factors include deregulation, increased competition and technological progress that together have resulted in network and service providers focusing on the discovery of a so-called ‘killer application’ to help them regain their competitive edge. However, there has being much confusion over what exactly is this next ‘killer application’ which can be demonstrated by various research reports. Instead, network and service providers should be focused on the creation of a scalable service provisioning platform where a variety of services can be dynamically composed on the fly depending on mobile subscriber service demands. In addition, mobile users within this new era of telecommunications have higher expectations with what they expect from their service providers with regards to their Quality of Service (QoS) and price requirements. However, the current subscription model employed with telecommunications prohibits a true and open market. The Telecommunication Service Exchange (TSE) the subject of which is the discussion of this thesis is an exchange based marketplace where mobile users can purchase services on a per request basis outside of their subscription contracts allowing them to exert their bargaining power in the Business-to-Customer (B2C) market, while adopting a Service Oriented Architectural (SPA) approach to service provisioning in the Business-to-Business (B2B) market allowing service providers to become more flexible in their business processes. Key findings of the work presented in this thesis demonstrate that such transactions in both markets take place in a reasonable time where mobile users can purchase services in approx 10 seconds outside of their subscription contracts and that B2B plan selection for use in a business process takes place in 1 to 4 seconds demonstrating the TSE feasibility and viability as a service provisioning platform for Beyond 3G (B3G).
Acknowledgements

And so it truly is with great pleasure that I write this final page of my thesis, a page where I can thank and acknowledge the people who have helped throughout my PhD.

First of all, I would like to thank Dr. Dirk Pesch, my project supervisor over the past three years. Without Dirk I don't believe I would ever have considered doing a master's not to mind a PhD degree. His unwavering encouragement, enthusiasm and technical guidance have also inspired me to pursuing my own career in research.

I would also like to thank my external and internal examiners, Dr. Dave Lewis (Trinity College, Dublin) and Dr. Martin Klepal. I would like to thank you both for a surprisingly enjoyable viva where your most useful comments and suggestions have allowed me to provide this final document.

Within Cork Institute of Technology (CIT) there are many people who have also supported me throughout the years both at a personal and academic level - to them I extend my thanks in maintaining my personal sanity throughout the process.

I would like to thank my parents - Flor and Angela and my sisters and brothers. In particular my mother, Angela, who has always being there at the exact moments when I have waived in my resolve to complete this PhD and to her I owe a most important debt of gratitude.

Finally I would like to extend my thanks to Ciaran and our daughter Avril. Avril has being the source of my determination over the past 5 years and to her I dedicate this thesis.
# Table of Contents

## Chapter 1

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Strategic Vision of Deregulation</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Mobile users want more than basic reliable services</td>
<td>2</td>
</tr>
<tr>
<td>1.1.3</td>
<td>Inefficiency of competition for the mobile user</td>
<td>5</td>
</tr>
<tr>
<td>1.2</td>
<td>Contribution of thesis</td>
<td>6</td>
</tr>
<tr>
<td>1.3</td>
<td>Goals of Research</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Synopsis</td>
<td>11</td>
</tr>
</tbody>
</table>

## Chapter 2

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>State of the Art</td>
<td>13</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Current Service Provisioning Approaches</td>
<td>14</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Evolution of Mobile Communications</td>
<td>14</td>
</tr>
<tr>
<td>2.1.2.1</td>
<td>Service Provisioning Approaches</td>
<td>15</td>
</tr>
<tr>
<td>2.1.2.2</td>
<td>Intelligent Networks and IN</td>
<td>15</td>
</tr>
<tr>
<td>2.1.2.3</td>
<td>Universal Mobile Telecommunication System</td>
<td>16</td>
</tr>
<tr>
<td>2.1.2.4</td>
<td>Application Servers and Gateway Interfaces</td>
<td>17</td>
</tr>
<tr>
<td>2.2</td>
<td>Marketplaces</td>
<td>20</td>
</tr>
<tr>
<td>2.3</td>
<td>QoS based service composition/allocation</td>
<td>28</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Mobile users selection of services</td>
<td>28</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Service Providers selection of services based on QoS</td>
<td>29</td>
</tr>
<tr>
<td>2.4</td>
<td>Summary and Motivation for the TSE</td>
<td>37</td>
</tr>
</tbody>
</table>

## Chapter 3

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>TSE Design</td>
<td>41</td>
</tr>
<tr>
<td>3.2</td>
<td>TSE Features</td>
<td>42</td>
</tr>
<tr>
<td>3.3</td>
<td>Service Oriented Architecture</td>
<td>43</td>
</tr>
<tr>
<td>3.4</td>
<td>Stateful Web Services</td>
<td>45</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Agents</td>
<td>49</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Applications of Agents</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Agents, Semantics and Ontologies</td>
<td>50</td>
</tr>
</tbody>
</table>
3.5 Automated Negotiation.......................................................... 52
  3.5.1 Multi-Attribute Auctions.................................................. 56
3.6 Preference Elicitation............................................................ 59
3.7 TSE Architecture.................................................................... 62
  3.7.1 Buyer User Agent (BUA).................................................... 64
  3.7.2 Market Interface Agent (MIA)............................................. 65
  3.7.3 Trusted Intermediary Agent (TIA)....................................... 65
  3.7.4 Service Provider Agent (SPA)............................................ 66
  3.7.5 Network Operator Agent (NOA)......................................... 66
  3.7.6 Better Business Bureau Agent (BBBA)................................. 66
  3.7.7 WS-Resource Integration Gateway Agent (WSRIGA)............ 67
3.8 TSE Negotiation Model........................................................ 69
3.9 Conclusion.............................................................................. 74

Chapter 4
TSE Operation............................................................................ 75
  4.1 TSE Mechanism .................................................................... 76
    4.1.1 Buyers Preferences........................................................... 83
    4.1.2 SPA Strategies............................................................... 86
  4.2 Initialisation of the TSE.......................................................... 88
    4.2.1 Create and Register Service Description.......................... 88
    4.2.2 Subscribe to services...................................................... 90
  4.3 B2C Market Operation............................................................ 93
    4.3.1 Initial Buyers Request..................................................... 93
    4.3.2 TIA Assignment............................................................. 96
    4.3.3 Call for Proposals.......................................................... 98
    4.3.4 Calculate and Return Bid................................................. 100
    4.3.5 Winner Determination................................................... 105
    4.3.6 Connection Establishment............................................... 108
    4.3.6.1 SIP Based IMS Procedures for Real Time Services........ 108
    4.3.6.2 Mobile Originating Call Setup....................................... 110
    4.3.7 Connection Termination.................................................. 114
  4.4 B2B Market Operation............................................................ 115
    4.4.1 Initial Service Providers Request..................................... 116
### 4.4.2 Call for Proposals ................................................................. 117
### 4.4.3 Calculate and Return Bid .................................................... 118
### 4.4.4 Process Generation ............................................................. 118
### 4.4.5 Subscribe to new Business Process ..................................... 119
### 4.4.6 Service Execution and Availability ..................................... 122
### 4.5 Conclusion ........................................................................... 126

#### Chapter 5 TSE Implementation and Analysis .......................... 128
#### 5.1 Agent Platforms ................................................................. 130
##### 5.1.1 JADE Architecture and TSE Ontology ........................... 135
#### 5.2 Web service Integration ....................................................... 139
#### 5.3 JADE Messaging and Measurement Issues .......................... 141
#### 5.4 DF Scalability ................................................................. 144
#### 5.5 WSRIG Performance ......................................................... 150
#### 5.6 Marketplace Performance Evaluation ................................. 156
##### 5.6.1 B2C Marketplace Evaluation ....................................... 159
##### 5.6.2 B2B Marketplace Evaluation ....................................... 162
#### 5.7 UMTS Performance Evaluation .......................................... 168
##### 5.7.1 SigSim Architecture and Parameters .............................. 169
##### 5.7.2 SIP Performance Evaluation ....................................... 170
##### 5.7.3 Performance Analysis .................................................. 172
#### 5.8 Conclusion ........................................................................... 178

#### Chapter 6 Conclusions and Future Work .................................. 181
#### 6.1 Conclusion ........................................................................... 182
#### 6.2 Future Work ......................................................................... 188

#### Appendix A Session Initiation Protocol .................................... 191
#### A.1 SIP Architectural Components ............................................ 192
#### A.2 SIP Messages ....................................................................... 193
#### A.3 SIP Java Development ....................................................... 195
### Appendix B

**Web services and Grid Computing**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.1</td>
<td>Web Services</td>
<td>196</td>
</tr>
<tr>
<td>B.1.1</td>
<td>Web Services Standardisation</td>
<td>196</td>
</tr>
<tr>
<td>B.1.2</td>
<td>The Basic Profile</td>
<td>197</td>
</tr>
<tr>
<td>B.1.3</td>
<td>Event Driven Architecture (EDA)</td>
<td>200</td>
</tr>
<tr>
<td>B.1.4</td>
<td>OWL-S and WSMO</td>
<td>202</td>
</tr>
<tr>
<td>B.2</td>
<td>Grid Computing</td>
<td>203</td>
</tr>
<tr>
<td>B.2.1</td>
<td>Introduction</td>
<td>203</td>
</tr>
<tr>
<td>B.2.2</td>
<td>Standardisation Bodies</td>
<td>204</td>
</tr>
<tr>
<td>B.2.3</td>
<td>Standardization Efforts</td>
<td>205</td>
</tr>
<tr>
<td>B.2.4</td>
<td>The Semantic Grid</td>
<td>208</td>
</tr>
</tbody>
</table>

### Appendix C

**Agents and Negotiation**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.1</td>
<td>Agents</td>
<td>213</td>
</tr>
<tr>
<td>C.1.1</td>
<td>Agent Communication Languages</td>
<td>213</td>
</tr>
<tr>
<td>C.1.2</td>
<td>Theories of Agency</td>
<td>216</td>
</tr>
<tr>
<td>C.1.3</td>
<td>SPA Agent Definition File</td>
<td>218</td>
</tr>
<tr>
<td>C.2</td>
<td>Negotiation Analysis</td>
<td>223</td>
</tr>
<tr>
<td>C.2.1</td>
<td>Game Theory</td>
<td>224</td>
</tr>
<tr>
<td>C.2.2</td>
<td>Mechanism Design</td>
<td>226</td>
</tr>
<tr>
<td>C.2.3</td>
<td>Terms and extensions to auction proposals</td>
<td>228</td>
</tr>
</tbody>
</table>

### Appendix D

**IMS & SigSim**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.1</td>
<td>IP Multimedia Subsystem (IMS)</td>
<td>231</td>
</tr>
<tr>
<td>D.2</td>
<td>Simulation Inputs</td>
<td>233</td>
</tr>
<tr>
<td>D.3</td>
<td>Network Element Implementation</td>
<td>236</td>
</tr>
</tbody>
</table>

### Bibliography

| Page | |
|------||
| 239  | |
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1-1</td>
<td>Cost and Benefits of Niche Services</td>
<td>4</td>
</tr>
<tr>
<td>Figure 2-1</td>
<td>Three Models of Competitive Negotiation in Marketplaces</td>
<td>21</td>
</tr>
<tr>
<td>Figure 2-2</td>
<td>CBB Model</td>
<td>22</td>
</tr>
<tr>
<td>Figure 2-3</td>
<td>BTT Model</td>
<td>22</td>
</tr>
<tr>
<td>Figure 2-4</td>
<td>SOA Architecture</td>
<td>29</td>
</tr>
<tr>
<td>Figure 2-5</td>
<td>SOA Collaborations</td>
<td>30</td>
</tr>
<tr>
<td>Figure 2-6</td>
<td>Orchestration versus Choreography Web services</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2-7</td>
<td>A Framework for Service Composition</td>
<td>35</td>
</tr>
<tr>
<td>Figure 3-1</td>
<td>TSE Market Architecture</td>
<td>62</td>
</tr>
<tr>
<td>Figure 3-2</td>
<td>TSE Internal Architecture</td>
<td>64</td>
</tr>
<tr>
<td>Figure 3-3</td>
<td>Integration Architecture</td>
<td>68</td>
</tr>
<tr>
<td>Figure 3-7</td>
<td>Auction types where quality is an issue</td>
<td>70</td>
</tr>
<tr>
<td>Figure 4-1</td>
<td>Preference Elicitation GUI</td>
<td>77</td>
</tr>
<tr>
<td>Figure 4-2</td>
<td>Effect of cost on Quality Offered</td>
<td>85</td>
</tr>
<tr>
<td>Figure 4-3</td>
<td>Effect of weights on Quality Offered</td>
<td>85</td>
</tr>
<tr>
<td>Figure 4-4</td>
<td>Effect of number of service providers (n) on price</td>
<td>86</td>
</tr>
<tr>
<td>Figure 4-5</td>
<td>Effect of rising cost on price</td>
<td>87</td>
</tr>
<tr>
<td>Figure 4-6</td>
<td>Effect of rising weights on price</td>
<td>87</td>
</tr>
<tr>
<td>Figure 4-7</td>
<td>Preliminary Operations</td>
<td>88</td>
</tr>
<tr>
<td>Figure 4-8</td>
<td>Belief Base of SPA</td>
<td>90</td>
</tr>
<tr>
<td>Figure 4-9</td>
<td>Subscribe to service</td>
<td>91</td>
</tr>
<tr>
<td>Figure 4-10</td>
<td>Initial Buyer Request Procedures</td>
<td>93</td>
</tr>
<tr>
<td>Figure 4-11</td>
<td>Session level Attributes</td>
<td>95</td>
</tr>
<tr>
<td>Figure 4-12</td>
<td>TIA Assignment</td>
<td>96</td>
</tr>
<tr>
<td>Figure 4-13</td>
<td>Call for Proposals</td>
<td>98</td>
</tr>
<tr>
<td>Figure 4-14</td>
<td>QoS Management Functions</td>
<td>101</td>
</tr>
<tr>
<td>Figure 4-15</td>
<td>Network Operations for calculating and Return Bid</td>
<td>103</td>
</tr>
<tr>
<td>Figure 4-16</td>
<td>TSE Calculate and Return Bid Procedures</td>
<td>104</td>
</tr>
<tr>
<td>Figure 4-17</td>
<td>Winner Determination</td>
<td>105</td>
</tr>
<tr>
<td>Figure 4-18</td>
<td>FIPA ACL Message Sequences for B2C Market</td>
<td>107</td>
</tr>
<tr>
<td>Figure 4-19</td>
<td>BUA Connection Establishment Procedures</td>
<td>108</td>
</tr>
<tr>
<td>Figure 4-20</td>
<td>MO Call Signalling with QoS Guarantees</td>
<td>113</td>
</tr>
<tr>
<td>Figure 4-21</td>
<td>Connection Termination</td>
<td>114</td>
</tr>
<tr>
<td>Figure 4-22</td>
<td>B2B ACL Message Sequences</td>
<td>121</td>
</tr>
<tr>
<td>Figure 4-23</td>
<td>Enterprise Service Bus in Network Operators Network</td>
<td>124</td>
</tr>
<tr>
<td>Figure 5-1</td>
<td>Jadex Abstract Architecture</td>
<td>132</td>
</tr>
<tr>
<td>Figure 5-2</td>
<td>XML ADF and Plans for Agents</td>
<td>132</td>
</tr>
<tr>
<td>Figure 5-3</td>
<td>Trace Exploration Graph SPA1</td>
<td>134</td>
</tr>
<tr>
<td>Figure 5-4</td>
<td>Reference Architecture of a FIPA Agent Platform</td>
<td>135</td>
</tr>
<tr>
<td>Figure 5-5</td>
<td>The Content Reference Model</td>
<td>137</td>
</tr>
<tr>
<td>Figure 5-6</td>
<td>Jade Messaging Architecture</td>
<td>142</td>
</tr>
<tr>
<td>Figure 5-7</td>
<td>Number of active agent during the considered time intervals</td>
<td>143</td>
</tr>
<tr>
<td>Figure 5-8</td>
<td>Register Service</td>
<td>144</td>
</tr>
<tr>
<td>Figure 5-9</td>
<td>Group Search Operation</td>
<td>145</td>
</tr>
<tr>
<td>Figure 5-10</td>
<td>Individual Search Operation</td>
<td>146</td>
</tr>
<tr>
<td>Figure 5-11</td>
<td>Modify Service Descriptions</td>
<td>147</td>
</tr>
<tr>
<td>Figure 5-12</td>
<td>Individual Modify Operations</td>
<td>147</td>
</tr>
<tr>
<td>Figure 5-13</td>
<td>Deregister Service Descriptions</td>
<td>148</td>
</tr>
<tr>
<td>Figure 5-14</td>
<td>Search Operation with search constraints</td>
<td>149</td>
</tr>
<tr>
<td>Figure 5-15</td>
<td>TSE, WSRIG and Web services</td>
<td>150</td>
</tr>
<tr>
<td>Figure 5-16</td>
<td>Agent exposed as Web service endpoint</td>
<td>152</td>
</tr>
<tr>
<td>Figure 5-17</td>
<td>10 Agents exposed as Web service endpoint</td>
<td>153</td>
</tr>
<tr>
<td>Figure 5-18</td>
<td>50 Agents exposed as Web service endpoint</td>
<td>153</td>
</tr>
<tr>
<td>Figure 5-19</td>
<td>100 Agents exposed as Web service endpoint</td>
<td>154</td>
</tr>
<tr>
<td>Figure 5-20</td>
<td>WSIG Processing Request/Subscribe</td>
<td>155</td>
</tr>
<tr>
<td>Figure 5-21</td>
<td>RTT for Request/Subscribe Procedure</td>
<td>155</td>
</tr>
<tr>
<td>Figure 5-22</td>
<td>TSE B2B and B2C markets including interoperable software components &amp; UMTS Architecture</td>
<td>156</td>
</tr>
<tr>
<td>Figure 5-23</td>
<td>B2C with 2 SPAs</td>
<td>159</td>
</tr>
</tbody>
</table>
Figure 5-24  B2C with 4 SPAs.......................................................... 160
Figure 5-25  B2C with 6 SPAs.......................................................... 160
Figure 5-26  RoundTrip Time for B2C Market................................. 161
Figure 5-27  2 Services 2 SPAs....................................................... 163
Figure 5-28  2 Services 4 SPAs....................................................... 163
Figure 5-29  4 Services 2 SPAs....................................................... 164
Figure 5-30  4 Services 4 SPAs....................................................... 165
Figure 5-31  6 Services 2 SPAs....................................................... 166
Figure 5-32  6 Services 4 SPAs....................................................... 166
Figure 5-33  RTT for B2B market for 2 SPAs................................. 167
Figure 5-34  RTT for B2B market for 4 SPAs................................. 167
Figure 5-35  TSE and General UMTS Architecture.......................... 168
Figure 5-36  SigSim Structure.......................................................... 169
Figure 5-33  RTT for B2C Market including NOA QoS Management
               Procedures................................................................. 176

App B F1  Web service Technology Stack........................................ 198
App B F2  WS-Addressing message information headers.................... 200
App B F3  JAIN SLEE Architecture................................................ 201
App B F4  Top Level of the service ontology.................................... 203
App B F5  Convergence of Web service and OGSI/WS-RF Services...... 210
App C F1  Progression of Agent Communication Language................ 214
App D F1  IMS Functional Planes.................................................... 232
List of Tables

Table 2-1 Classification of Controlled Multi-Agent E-Marketplaces...... 21
Table 2-2 Features of Multi-Agent E-Marketplaces with Negotiation
Model A........................................................................................................ 25
Table 2-3 B2C Mediation activities of eMarketplaces with negotiation
Model A........................................................................................................ 25
Table 2-4 The Digital Marketplace (DMP) Characteristics ....................... 26
Table 2-5 CBB Model for DMP.................................................................... 27

Table 3-1 Agent and Web service communication languages and
protocols......................................................................................................... 67
Table 3-2 3GPP End User expectations of RT services............................... 71

Table 4-1 Matrix of Pair Wise Comparisons................................................ 77
Table 4-2 Eigenvector.................................................................................... 77
Table 4-3 SPA Bids....................................................................................... 105
Table 4-4 SPA Bids....................................................................................... 119

Table 5-1 Agent Platforms........................................................................... 130
Table 5-2 TSE Ontology............................................................................... 137
Table 5-3 Hardware and Software testbed configuration......................... 142
Table 5-4 Interface Bandwidth..................................................................... 171
Table 5-5 SIP Session Setup Message Sizes............................................. 172
Table 5-6 PDP Activation/Deactivation Times............................................ 173
Table 5-7 NOA and Network Procedures for QoS Management
Functions...................................................................................................... 173
Table 5-8 Compressed SIP Session Setup Message Sizes......................... 173
Table 5-9 Mobile Originated SIP call setup delay 3.4kb/s bearer.............. 174
Table 5-10 Mobile Terminated SIP call setup delay 3.4kb/s bearer............ 174
Table 5-11 Mobile Originated SIP call setup delay 64kb/s bearer............. 175
Table 5-12 Mobile Terminated SIP call setup delay 64kb/s bearer.......... 175
<table>
<thead>
<tr>
<th>Table 5-13</th>
<th>Overall timing for purchasing a voice call on a per request basis</th>
<th>176</th>
</tr>
</thead>
<tbody>
<tr>
<td>App A T1</td>
<td>SIP Requests</td>
<td>193</td>
</tr>
<tr>
<td>App A T2</td>
<td>Response Codes</td>
<td>194</td>
</tr>
<tr>
<td>App A T3</td>
<td>JCP SIP Programming API</td>
<td>195</td>
</tr>
<tr>
<td>App B T1</td>
<td>BP 1.0 Standards</td>
<td>198</td>
</tr>
<tr>
<td>App B T2</td>
<td>Standardisation efforts of the Grid</td>
<td>206</td>
</tr>
<tr>
<td>App B T3</td>
<td>OGSI Refactoring</td>
<td>207</td>
</tr>
<tr>
<td>App B T4</td>
<td>The Grid v’s The Semantic Web</td>
<td>210</td>
</tr>
<tr>
<td>App C T1</td>
<td>FIPA ACL Interaction Protocols</td>
<td>215</td>
</tr>
<tr>
<td>App D T1</td>
<td>Network Configuration Data</td>
<td>233</td>
</tr>
<tr>
<td>App D T2</td>
<td>User Profile Data</td>
<td>234</td>
</tr>
<tr>
<td>App D T3</td>
<td>Service-related Data</td>
<td>235</td>
</tr>
<tr>
<td>App D T4</td>
<td>Implemented Network Element Assumptions</td>
<td>237</td>
</tr>
<tr>
<td>App D T5</td>
<td>SGSN Turn-around times</td>
<td>238</td>
</tr>
<tr>
<td>App D T6</td>
<td>GGSN Turn-around times</td>
<td>238</td>
</tr>
</tbody>
</table>
### Abbreviations & Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACL</td>
<td>Agent Communication Language</td>
</tr>
<tr>
<td>ACC</td>
<td>Agent Communication Channel</td>
</tr>
<tr>
<td>ADF</td>
<td>Agent Definition File</td>
</tr>
<tr>
<td>ADK</td>
<td>Agent Development Kit</td>
</tr>
<tr>
<td>AHP</td>
<td>Analytical Hierarchical Process (AHP)</td>
</tr>
<tr>
<td>AID</td>
<td>Agent Identifier</td>
</tr>
<tr>
<td>AMPS</td>
<td>Advanced Mobile Telephone System</td>
</tr>
<tr>
<td>AMS</td>
<td>Agent Management System</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>AS</td>
<td>Application Server</td>
</tr>
<tr>
<td>B2B</td>
<td>Business-to-Business</td>
</tr>
<tr>
<td>B2C</td>
<td>Business-to-Consumer</td>
</tr>
<tr>
<td>B3G</td>
<td>Beyond 3G</td>
</tr>
<tr>
<td>BBBA</td>
<td>Better Business Bureau Agent</td>
</tr>
<tr>
<td>BBT</td>
<td>Business Buyer Model</td>
</tr>
<tr>
<td>BDI</td>
<td>Belief Desire Intention</td>
</tr>
<tr>
<td>BP</td>
<td>Basic Profile</td>
</tr>
<tr>
<td>BPEL</td>
<td>Business Process Execution Language</td>
</tr>
<tr>
<td>C2B</td>
<td>Consumer-to-Business</td>
</tr>
<tr>
<td>C2C</td>
<td>Consumer-to-Consumer</td>
</tr>
<tr>
<td>CAMEL</td>
<td>Customized Application for Mobile Enhanced Logic</td>
</tr>
<tr>
<td>CB</td>
<td>Citizen Band</td>
</tr>
<tr>
<td>CBB</td>
<td>Customer Buying Model</td>
</tr>
<tr>
<td>CDA</td>
<td>Continuous Double Auction</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code Division Multiple Access</td>
</tr>
<tr>
<td>CCL</td>
<td>Constraint Choice Language</td>
</tr>
<tr>
<td>CN</td>
<td>Core Network</td>
</tr>
<tr>
<td>CFP</td>
<td>Call For Proposals</td>
</tr>
<tr>
<td>CSSL</td>
<td>Composite Service Specification Language</td>
</tr>
<tr>
<td>CSTA</td>
<td>Computer Supported Telecommunication Applications</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>DFAD</td>
<td>Directory Facilitator Agent Descriptions</td>
</tr>
<tr>
<td>DMP</td>
<td>Digital Market Place</td>
</tr>
<tr>
<td>DF</td>
<td>Directory Facilitator</td>
</tr>
<tr>
<td>ECMA</td>
<td>European Computer Manufacturers Association</td>
</tr>
<tr>
<td>eMarketplace</td>
<td>Electronic Marketplace</td>
</tr>
<tr>
<td>eCommerce</td>
<td>electronic Commerce</td>
</tr>
<tr>
<td>EDA</td>
<td>Event Driven Architecture</td>
</tr>
<tr>
<td>EDGE</td>
<td>Enhanced Data Rate for GSM Evolution</td>
</tr>
<tr>
<td>eMarketplace</td>
<td>electronic Marketplace</td>
</tr>
<tr>
<td>FPSB</td>
<td>First Price Sealed Bid</td>
</tr>
<tr>
<td>KIF</td>
<td>Knowledge Interchange Format</td>
</tr>
<tr>
<td>GPS</td>
<td>General Positioning System</td>
</tr>
<tr>
<td>GGF</td>
<td>Global Grid Forum</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile communications</td>
</tr>
<tr>
<td>HLR</td>
<td>Home Location Register</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>HSS</td>
<td>Home Subscriber Server</td>
</tr>
<tr>
<td>I-CSCF</td>
<td>Interrogating CSCF</td>
</tr>
<tr>
<td>IMS</td>
<td>IP Multimedia Subsystem</td>
</tr>
<tr>
<td>IN</td>
<td>Intelligent Networks</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>JADE</td>
<td>Java Agent Development Framework</td>
</tr>
<tr>
<td>JDK</td>
<td>Java Development Kit</td>
</tr>
<tr>
<td>JMX</td>
<td>Java Management Extension</td>
</tr>
<tr>
<td>JVM</td>
<td>Java Virtual Machine</td>
</tr>
<tr>
<td>MAGMA</td>
<td>Minnesota AGent Marketplace Architecture</td>
</tr>
<tr>
<td>MAGNET</td>
<td>Multi-Agent NEgotiation Testbed</td>
</tr>
<tr>
<td>MAUT</td>
<td>Multi-Attribute Utility Theory</td>
</tr>
<tr>
<td>MIAMAP</td>
<td>MIAMI Marketplace</td>
</tr>
<tr>
<td>MIDP</td>
<td>Mobile Information Device Profile</td>
</tr>
<tr>
<td>MGCF</td>
<td>Media Gateway Control Function</td>
</tr>
<tr>
<td>MGW</td>
<td>Media Gateway</td>
</tr>
<tr>
<td>MOWS</td>
<td>Management Of Web Services</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MSC</td>
<td>Mobile Switching Centre</td>
</tr>
<tr>
<td>MUWS</td>
<td>Management Using Web Services</td>
</tr>
<tr>
<td>WSDM</td>
<td>Web services Distributed Management</td>
</tr>
<tr>
<td>MVNO</td>
<td>Mobile Virtual Network Operator</td>
</tr>
<tr>
<td>NGN</td>
<td>Next Generation Network</td>
</tr>
<tr>
<td>NI</td>
<td>Network Initiated</td>
</tr>
<tr>
<td>NOA</td>
<td>Network Operator Agent</td>
</tr>
<tr>
<td>OAA</td>
<td>Open Agent Architecture</td>
</tr>
<tr>
<td>OASIS</td>
<td>Organization for the Advancement of Structured Information Standards</td>
</tr>
<tr>
<td>OGSA</td>
<td>Open Grid Service Architecture</td>
</tr>
<tr>
<td>OGSI</td>
<td>Grid Services Infrastructure</td>
</tr>
<tr>
<td>OIL</td>
<td>Ontology Inference Layer</td>
</tr>
<tr>
<td>OMA</td>
<td>Open Mobile Alliance</td>
</tr>
<tr>
<td>OSA</td>
<td>Open Service Access</td>
</tr>
<tr>
<td>OWL</td>
<td>Web Ontology Language</td>
</tr>
<tr>
<td>OWL-S</td>
<td>Semantic Web Ontology Language</td>
</tr>
<tr>
<td>P-CSCF</td>
<td>Proxy CSCF</td>
</tr>
<tr>
<td>PDDL</td>
<td>Planning Domain Definition Language</td>
</tr>
<tr>
<td>PDA</td>
<td>Personal Digital Assistant</td>
</tr>
<tr>
<td>PEP</td>
<td>Policy Enforcement Point</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework</td>
</tr>
<tr>
<td>RDF (S)</td>
<td>RDF (Schema)</td>
</tr>
<tr>
<td>ROHC</td>
<td>Robust Header Compression</td>
</tr>
<tr>
<td>RMI</td>
<td>Remote Method Invocation</td>
</tr>
<tr>
<td>SOA</td>
<td>Service Oriented Architecture</td>
</tr>
<tr>
<td>SMTP</td>
<td>Simple Mail Transfer Protocol</td>
</tr>
<tr>
<td>SigSim</td>
<td>Signalling Simulator</td>
</tr>
<tr>
<td>SCF</td>
<td>Service Capability Feature</td>
</tr>
<tr>
<td>SCS</td>
<td>Service Capability Server</td>
</tr>
<tr>
<td>S-CSCF</td>
<td>Serving CSCF</td>
</tr>
<tr>
<td>SDP</td>
<td>Session Description Protocol</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>SIP</td>
<td>Session Initiation Protocol</td>
</tr>
<tr>
<td>SLA</td>
<td>Service Level Agreements</td>
</tr>
<tr>
<td>SLEE</td>
<td>Service Logic Execution Environment</td>
</tr>
<tr>
<td>SMART</td>
<td>Simple Multi-Attribute Utility Theory</td>
</tr>
<tr>
<td>SOA</td>
<td>Service Oriented Architecture</td>
</tr>
<tr>
<td>SOAP</td>
<td>Simple Object Access Protocol</td>
</tr>
<tr>
<td>SPA</td>
<td>Service Provider Agent</td>
</tr>
<tr>
<td>SPSB</td>
<td>Second Price Sealed Bid</td>
</tr>
<tr>
<td>SWRL</td>
<td>Semantic Web Rule Language</td>
</tr>
<tr>
<td>TACS</td>
<td>Total Access Communication System</td>
</tr>
<tr>
<td>TCCB</td>
<td>Text-based Compression using Cache and Blank</td>
</tr>
<tr>
<td>TIA</td>
<td>Trusted Intermediary Agent</td>
</tr>
<tr>
<td>TSE</td>
<td>Telecommunication Service Exchange</td>
</tr>
<tr>
<td>UA</td>
<td>User Agent</td>
</tr>
<tr>
<td>UDDI</td>
<td>Universal Description Discovery Integration</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunication System</td>
</tr>
<tr>
<td>URI</td>
<td>Uniform Resource Identifier</td>
</tr>
<tr>
<td>VO</td>
<td>Virtual Organisation</td>
</tr>
<tr>
<td>VoIP</td>
<td>Voice over IP</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>W-CDMA</td>
<td>Wideband Code Division Multiple Access</td>
</tr>
<tr>
<td>WS-CDL</td>
<td>Web Service Choreography Description Language</td>
</tr>
<tr>
<td>WSDL</td>
<td>Web Service Description Language</td>
</tr>
<tr>
<td>WSDM</td>
<td>Web services Distributed Management</td>
</tr>
<tr>
<td>WSRIG</td>
<td>Web Service Resource Integration Gateway</td>
</tr>
<tr>
<td>WSRIGA</td>
<td>Web Service Resource Integration Gateway Agent</td>
</tr>
<tr>
<td>WS-I</td>
<td>Web Service Interoperability Organization</td>
</tr>
<tr>
<td>WSMF</td>
<td>Web Service Modelling Framework</td>
</tr>
<tr>
<td>WSMO</td>
<td>Web Service Modelling Ontology</td>
</tr>
<tr>
<td>WS RF</td>
<td>WS-Resource Framework</td>
</tr>
<tr>
<td>WSN</td>
<td>Web Service Notification</td>
</tr>
<tr>
<td>WWRF</td>
<td>Wireless World Research Forum</td>
</tr>
<tr>
<td>3GPP</td>
<td>3rd Generation Partnership Project</td>
</tr>
</tbody>
</table>
Variables Used

$q_u$  
quantity of a particular service instance

$N$  
number of sellers in the market

$n$  
number of bidders

$Q$  
quantity of units sold

$p$  
price

$b$  
bid

$BU$  
Buyer

$s$  
seller

$S$  
scoring function

$q_p$  
value for quality attribute performance

$q_A$  
value for quality attribute availability

$q_R$  
value for quality attribute rating

$q_T$  
value for quality throughput rating

$q_i$  
all the quality attributes i.e. $q_p, q_A, q_R, q_T$

$w_p$  
weight for quality attribute performance

$w_A$  
weight for quality attribute availability

$w_T$  
weight for quality attribute throughput

$w_R$  
weight for quality attribute rating
weights for all the quality attributes i.e. \( w_p, w_A, w_T, w_R \)

cost for producing the quality attribute performance

cost for producing the quality attribute availability

cost for producing the quality attribute throughput

cost for producing the quality attribute rating

cost for all quality attribute for service provider i i.e. \( c_p, c_A, c_T, c_R \)

private cost parameter

value

bid containing price and quality specification

utility

levels
To date, the work presented in this thesis has produced the following publications:


Chapter 1

Introduction

1.1 Strategic Vision of Deregulation

During the past few decades the field of telecommunications has been subject to a continuous evolution. Kridel [1] associates this evolution to three interrelated phenomena: shifts in regulations, increased competition and technological progress.

In the past, state ownership of critical resources such as electricity, gas and telecommunications, created monopolies in these areas throughout the world. However, deliberate government and regulatory policies in the US and Europe as well as technological advances and globalization have opened up markets and have given new businesses the right to compete in the telecommunication markets. FCC and Ofcom are examples of independent regulators in the US and UK, whose focus and goal is “to move from monopoly to competition” [2] across television, radio, telecommunications and wireless communications services. The effects of deregulation in the mobile telecommunications industry can be seen with the establishment of Mobile Virtual Network Operators (MVNOs), who “piggy-back” on an existing mobile network to offer competing mobile services. Virgin Mobile, which operates over the T-Mobile network in the UK and Sprint in the US, is the most successful MVNO globally. This success is demonstrated by the number of active subscribers of 4 and 4.88 million in the UK and US respectively as of April 2007\(^1\). While deregulation in the form of MVNOs has provided an additional form of competition within the sector, it has by no means addressed the strategic level objectives of the regulating authorities. Ofcom’s [3] strategic review of deregulation in telecommunications outlines the tactical direction of its activities in the future. Key aspects of this review are that:

1. Mobile users want much more than basic reliable telecom services. They also want choice, with rapid innovation and introduction of new services. Their assessment

\(^1\) Virgin Mobile, [http://about.virginmobile.com](http://about.virginmobile.com)
indicated that the most effective means of delivering this is through competition at the deepest level of infrastructure where competition will be effective and sustainable.

2. Competition within telecommunications cannot be effective unless consumers are able to make well-informed choices, and can switch easily between suppliers [3].

In order to provide a rationale for the importance of the work presented in this thesis it is necessary to provide a discussion of these strategic level objectives from the mobile user and service providers point of view.

1.1.1 Mobile users want much more than basic reliable telecom services

Voice telephony was the only available service provided by 1st Generation mobile communication systems and is still the main means of generating revenue in its 3rd Generation systems. The dominance of voice as the primary revenue generating service allowed service and network operators to adopt a walled garden approach to their networks, where Jaokar and Fish [5] define a walled garden is a "mechanism for an entity to restrict the user experience by confining the user to a specific region/space as defined by the entity", the rational behind this being that the user is served better and the service is more profitable for the provider i.e. Vodafone live.

However in the future it is envisioned by the UMTS forum that if one third of subscribers are on 3G networks by 2010, the cumulative revenue from 3G services is expected to be over one trillion dollars with 66% of these revenues coming from 3G enabled data services [6]. In such an environment mobile users will want much more than the basic services such as a voice call and service providers will have to provide these services to meet their demands, maintain their competitive edge and differentiate themselves from their competitors. To do this effectively service providers will have to adopt an open access policy to their networks, giving the mobile user unrestricted access to whatever content is available on the mobile web where the service or network operators charge purely on traffic i.e. i-mode, Meteor Stuff!, while at the same time giving themselves access to open source third party applications and services.

However this open access model to telecommunications has being subject to a continuous debate and is highlighted in [7] which states that "some carriers will balk at
the idea of exposing their networks with a standard-based gateway such as Parlay and Parlay-X’. This ‘balking’ has emerged due to a number of competitive factors where deregulation in the sector has added new players and consequently choices that mobile users have for business and personal communication. In addition, new technologies within the sector has allowed service providers to build and operate networks, providing local, long, global, and mobile voice, data, video, and Internet services to businesses and consumers. In addition, mobile network operators had to incur a significant investment cost with the purchase of their 3G licenses, with $700 billion being spent in 3G spectrum auctions in Europe alone [4]. Such competitive forces has made network operators protective over their existing client base where they believe that protecting their installed consumer base outweighs the potential revenue from new consumers, thereby adopting a ‘walled garden’ approach to their networks.

In addition to the advantages of service differentiation, openness in telecommunications will also insert ‘new blood’ into the industry where so called ‘grassroots developers’ outside the domain of telecommunications can foster new ideas and consequently unfretted innovation will emerge. In the past in other industries new ideas have frequently come from outside the industry, where for example the work of Josef Gregor Mendel, who was a priest, made a impact in the field of genetics through his observations on the diversity of pea plants.

From a business point of view however, developing innovative services is not as important as creating innovative services that mobile users are willing to pay for. This distinction was highlighted in the past with the videophone application that AT&T first developed, where since the early days of communications the industry was fascinated with combining voice and video into a single application. While its development was hailed a scientific achievement, the videophone application failed commercially despite several media campaigns. In order for service and network operators to protect themselves again from such commercial failures it will be not only necessary for them to adopt a open access policy to their networks, but also to change their current large scale, macro service delivery paradigm to focus on many more focused micro-services where the relationship between micro and macro services is shown in Figure 1-1.
Adopting this approach minimises risk where service providers can test the subscriber take up with micro services focused at niche markets, giving a pragmatic and practical proof of concept. Utilizing micro services also gives the option of reusability and creates a situation that even if the service context changes the micro-components could easily be transferred into a new context, again minimising risk and maximising revenue [9]. These micro-services can be enabled by the adoption of a Service Oriented Architecture (SOA) and Web services that enable them, where Channabasavaiah [10] defines a SOA as "an application architecture within which all functions are defined as independent services with well defined invokable interfaces which can be called in defined sequences to form business processes". A service in SOA can be defined as an "application function packaged as a reusable component for use in a business process". In essence, services within the SOA are self contained, modular, interoperable, loosely coupled, location transparent and composite entities. If such an approach was adapted to the development of the videophone application service providers would have created the application in a manner which facilitated the reuse of the service components and deployed the service to a particular segment of the market. Testing the subscriber take up within this niche market would then have highlighted at an early stage the mobile users disinterest in the service and would have facilitated the owning service provider to make the constituent components available to another service or application, enabling
the service provider to reuse the service, increasing its flexibility in its business processes, in addition to reducing the cost and risk of undertaking the project.

From a technical perspective, providing SOA within telecommunications has been in the past hindered by monolithic service architectures where services were delivered through vertically integrated service specific networks. If such a horizontal approach to service provisioning is adopted in telecommunications it is envisioned according to the Wireless World Research Forum\(^1\) (WWRF) [8] that this will allow innovative services to be created and deployed in a short time, addressing user needs, with third party interfaces allowing a chaining of expertise in service provisioning, enabling service and network operators to become more flexible and adaptable in their business processes.

In essence to fully unleash the mobile operators revenue potential and in meeting the demands on mobile users in the future with regards to innovative services, their walled garden approach to service provisioning will have to be questioned and operators will have to adopt SOA and Web services allowing them to become fully responsive to changing mobile user requirements and competition in the sector.

1.1.2 Inefficiency of competition for the mobile user

Adam Smith [11] observed that "Man is an animal that needs bargains: no other animal does this - no dog exchanges bones with another".

Within telecommunications at present, mobile users are typically tied to their provider via a long term contract lasting usually 12 months. At the end of their subscription contracts, mobile users are endowed with the choice of renewing their contract with their existing service provider or change their custom to a competitor. In order to assist mobile users in switching between service providers at the end of their subscription contracts, number portability was introduced within the European Union (EU) enabling subscribers of publicly available telephone services (including mobile services) to change their service provider whilst keeping their existing telephone number. Its purpose was to foster consumer choice and competition by enabling subscribers to

\(^1\) WWRF was established in that aims to develop a common global vision for future wireless research and standardisation with the aim providing information to the community to better understand what is relevant in research and thereby reduce risks in research. The WWRF is not a standardisation body but rather supports the 3GPP, 3GPP2, European Telecommunications Standards Institute (ETSI), UMTS, IETF and other relevant bodies relating to commercial and standardisation issues derived from the research work.
switch between providers without the costs and inconvenience of changing telephone number. However, this approach still does not address the subscription issue which lies at the core of a mobile user's inability to exert their bargaining power, where within their contract the mobile user cannot switch from one service provider to another to avail of special offers and services that the alternative service provider may be capable of offering. This causes an inefficiency of competition in telecommunications from the mobile user's perspective. However, allowing consumers to purchase services on a per request basis, while at the same time maintaining their contract with their chosen service providers however would provide more competition within the sector, and will force service providers to better serve the interests of users.

The basis on this proposition is led by Adam Smith's *invisible hand* argument, where it is argued that opening up a market within telecommunications will result in a globally efficient mechanism where the mobile user is able to choose freely what to buy, service providers are allowed to choose freely what to sell and how to produce it, which will in turn lead the market to settle on a product distribution and prices that will be beneficial to all members of the community. The concept of Adam Smith's invisible hand argument is used within the contribution of the work presented in this thesis.

1.1.3 Contribution of thesis

Looking at the strategic level objectives of Ofcom [3] it was aim of this thesis to design a service provisioning platform that can enable service providers to provide new services to the mobile user's quickly, while at the same time addressing the service providers competitive, technical and business constraints. In addition the service platform should address the inefficiency of competition from the mobile user's perspective as a result of the restrictions imposed on the mobile user with the current subscription model.

Bearing the above requirements in mind and using the concept of Adam Smiths Invisible *Hand* argument, the Telecommunication Service Exchange (TSE) [13, 14] is presented in this thesis as the core contribution, which is a service provisioning platform for Beyond 3G (B3G).

Within the contribution, using the principles of SOA and Web services would enable service providers to dynamically register services in order to sell them as interoperable
software components that encapsulate business functionalities available over the Internet. With these registered services, service providers are then endowed with the flexibility of choosing software components (i.e. represented as Web services) for a new service that they wish to offer within their network where they can then pay for this service on an as-needed basis. Dynamic service composition techniques drawn from the Artificial Intelligence (AI) domain can be employed to actually compose and execute the service. Newly composed services can then be offered within the network operator’s home network and can also be recomposed on the fly depending on the performance of the atomic service elements. These features increase the range of service providers and their respective services to collaborate with, enabling service and network operators to become more flexible and adaptable in their business processes, in addition to facilitating the reuse of existing software components thereby reducing the risk and cost of undertaking new projects.

As a result of the limitation imposed by subscription contracts, it is also the aim of the contribution to allow mobile users to purchase a service on a per request basis, thereby exerting their bargaining power, while at the same time maintaining their current subscription contracts with their chosen service provider. A key feature of the proposition resides in the ability of the various stakeholders in the system to dynamically select the service components or desired service, according to their price and QoS constraints. Using the approach outlined above will introduce more competition for the provision of telecommunication services where its approach is actually in line with the recommendations from Ofcom, who have openly stated that the “mobile market is not yet fully effective” due to lack of true competition in the telecommunications industry. Such an open access environment has also being advocated by Lewis et al [17] within telecommunications, where [17] outlines the Telecommunication Information Networking Architecture (TINA) business model that can be applied to such an open services market.

In order to ensure the integrity of the TSE goals in: open access; dynamic service composition and; mobile users ability to exert the bargaining power, the system would have to be controlled by a trusted third party or regulating authority such as Ofcom or FCC. In addition, to obtain willing participation from the various stakeholders operating within the TSE, its inherent features need to be propositioned in a manner where its
value for deployment outweighs its disadvantages of additional: signalling in the network operator’s network where a key component of this issues lies in the standardisation of such; and delay that the mobile user will incur for purchasing the service on a per request basis.

It is the purpose of this thesis therefore to demonstrate the viability and feasibility of the proposed TSE framework as a service provisioning platform for B3G. It does this by providing a critical assessment of the: integration issues that exist between the TSE and service provider’s home network and; additional signalling procedures that need to be employed in the network operator’s home network. This assessment is required to eliminate the technical concerns that a network operator may have in supporting the framework and to present such to the standardisation bodies for incorporating into technical specifications in the future. In addition a quantitative assessment addressing the question of ‘how long’ will regards to: purchasing a service on a per request basis from the mobile user’s perspective and; the dynamic collaboration procedures between service providers is addressed within the scope of this study. This is important as the mobile user and service provider will have the expectation of a reasonable delay in setup procedures and if this becomes excessive this delay may outweigh the advantage of giving the mobile user choice and service providers flexibility and adaptability in the TSE.

Based on the above discussion the contribution and claims of this thesis can be summarised as follows:

The core contribution of this thesis lies in the design and specification of the Telecommunication Service Exchange (TSE) framework which is a service provisioning platform for Beyond 3G (B3G). There are two main claims relating to this contribution.

The first claim is that deploying the TSE as a complementary approach to existing service provisioning methods within an network operators home network i.e. the IP Multimedia Subsystem (IMS), creates a flexible, adaptable and responsive environment by enabling:

- Mobile users purchase services on a per request basis
• Service providers to sell services where their interoperable business functionalities available over the Internet can form part of business collaborations

The second claim presented in this thesis is that the novel service provisioning platform developed as part of the contribution is a feasible and viable architecture where the acquisition of a service instance on behalf of a mobile user and the dynamic coalition procedures between service providers can take place in a reasonable time
1.2 Goals of Research

The main goals of this thesis can be summarized as follows:

1. Summarize current approaches to:
   a. Service provisioning in Universal Mobile Telecommunication System (UMTS) and their ability in addressing the strategic objectives of the regulating authorities
   b. Marketplaces where a comparative analysis is provided based on some attribute i.e. negotiation
   c. Quality of Service (QoS) based service selection from a mobile user and service providers perspective

2. Design a new service provisioning platform for Beyond 3G (B3G), which overcomes the limitations of existing approaches. Such a platform should utilize concepts and technologies which are suitable to enable competition and allow service providers to become more flexible and adaptable in their business processes

3. Implement the new service provisioning platform to prove the viability and feasibility of the designed framework

4. Conduct performance evaluation to provide a quantitative analysis addressing the question of 'how long' does the procedures takes from the mobile user and service providers perspective
1.3 Synopsis

Chapter 2 of this thesis provides a detailed state of the art into the current approaches to service provisioning in Universal Mobile Telecommunication System (UMTS) networks, marketplace approaches to service provisioning and contributions to component selection for use in a business process. Following the state of the art presented the motivation for developing a new service provisioning platform is outlined.

Chapter 3 commences by detailing the key features that the TSE must have to overcome the current approaches to service provisioning. Using these features a rationale for the design elements in the TSE is presented. The chapter provides a justification for the adoption of Service Oriented Architectures (SOA) within telecommunication where concrete evidence of the financial benefits of SOA is provided. The chapter outlines the management issues related to Web services with reference to stateful and persistent Web services. An overview of the main characteristics of agent technologies and a rationale for their use in the TSE is provided in this chapter, with the main agents and their functions that are used within the TSE are also outlined. Semantics and ontologies and their role within the TSE are detailed and how agent technology is an important technology for enabling the vision of semantic Web services. Automated negotiation techniques with particular reference to multi-attribute auctions is provided, in addition to the auction mechanism designed for the TSE.

Chapter 4 of this thesis looks at the internal operation procedures within the B2C and B2B markets of the TSE. It firstly outlines how the chosen preference elicitation technique, AHP, is adopted to the context of the TSE. The chapter then proceeds to outline how the auction mechanism outlined in Chapter 3 is used in the TSE operation and the Service Provider Agent (SPA) strategies computed using game theory. The B2C and B2B marketplace operation is also detailed where the signalling operations and protocols used to enable the TSEs functionality are provided. Finally, an overview of how the TSE can be used as a complementary approach to service provisioning with the IMS is also addressed.

Chapter 5 outlines the implementation and performance analysis aspects of the TSE. The main function of this chapter is to provide proof that the signalling procedures outlined in Chapter 4 are correct and that the claim is plausible to enable: the mobile
user purchase a service on a per request basis based on their QoS and price demands; and QoS driven atomic service selection for the purpose of dynamic service provisioning. Performance analysis of the Directory Facilitator (DF), Web Service Resource Integration Gateway (WSRIG), B2C and B2B markets are outlined, in addition to UMTS Release 5/6 network latency performance analysis aspects.

Chapter 6 provides a summary of the conclusions that can be drawn from the work presented and points out some directions for further research.

Appendix A to D provide additional information relating to the design and implementation of the TSE, namely it provides supplementary details on Session Initiation Protocol (SIP), Web services, Grid computing, Agents, negotiation analysis and simulation inputs that are used in the UMTS Signalling Simulator. Readers should refer to the Appendix in the event that a concept/technology is outside the scope of their research domain and may require supplementary tutorial information to aid their understanding.
In order to address the first goal of this thesis, it is necessary to provide a qualitative analysis into the current approaches to service provisioning and how their functionality does not achieve the regulators strategic objectives as outlined in Chapter 1. As the Telecommunication Service Exchange (TSE) is represented as an exchange based electronic Marketplace (eMarketplaces) a state of the art into marketplaces, with particular focus on marketplace approaches to service provisioning is also outlined. Also, since Quality of Service (QoS) and price constraints as specified by the mobile user is used to choose among a potentially large number of service providers capable of providing the service on a per request basis and; service providers to distinguish between other service providers and their respective services to collaborate with, a state of the art with regards to QoS based service selection and composition procedures are detailed. Finally, bearing the characteristics of: service provisioning platforms for Beyond 3G and; the strategic vision of the regulators in telecommunications in mind, limitations to the existing approaches are detailed to provide a concrete motivation for developing the Telecommunication Service Exchange (TSE).
2.1 Current Service Provisioning Approaches

This section commences with an overview on the evolution of mobile communications outlining a vision of mobile telecommunication systems into the future. The section then outlines the current approaches to service provisioning in UMTS detailing approaches such as functional planes, gateway interfaces and application servers.

2.1.1 Evolution of Mobile Communications

The first (analogue) cellular system became operational in Tokyo, Japan in 1979. The network was operated by NTT and utilized 600 duplex channels in the 800 MHz band. In 1981 the cellular era reached Europe. Nordic Mobile Telephone started operations in the 450 MHz band in Scandinavia. The Total Access Communication System (TACS) was launched in the United Kingdom in 1982, and Extended TACS was deployed in 1985. Germany introduced the C-450 cellular system in September 1985. By the end of the 1980s it was clear that first generation cellular systems were becoming obsolete, resulting in the evolution towards second-generation (digital) cellular systems in the early 1990s. Europe led the way by introducing the Global System for Mobile communications (GSM), providing a single unified standard in Europe.

The evolution in the United States was somewhat different. The first cellular mobile radio telephone system in the US was called the Advanced Mobile Telephone System (AMPS) which was specified by the Federal Communications Commission and major industry players like AT&T and Motorola. By the early 1990's analogue cellular had become so popular that the first generation analog systems couldn't keep up with demand. In the U.S three digital technologies were standardized and deployed: IS-136, IS-95 and Global System for Mobile Communications (GSM). GSM achieved global acceptance as it offered a rich selection of capabilities and features that provided real incremental revenues for operators. But times were changing again.

By the late 1990's the Internet had become pervasive and the wireless world looked to mobile data as the growth opportunity. Once again the industry undertook the task of defining new wireless systems – the 3rd Generation (3G) mobile system, called the Universal Mobile Telecommunication System (UMTS), which was based on packet data. Three new standards emerged: Code Division Multiple Access (CDMA2000),
Enhanced Data Rate for GSM Evolution (EDGE) and Wideband CDMA (W-CDMA). W-CDMA is the main third generation air interface in the world and has been deployed in the US, Europe and Asia, including Japan and Korea, in the same 2.1 GHz frequency [18].

However, over the years it has become evident that the future is much more complex than a new cellular radio system and related infrastructure. The next generation of mobile telecommunications systems, will depend on the services and applications provided. These future systems are expected to integrate the paradigms of traditional mobile telecommunication systems with the Internet Protocol (IP) suite and state-of-the-art software engineering methods. As a result of this it is envisioned that new paradigms will emerge [8]. On the basis of the third generation experience, future systems beyond 3rd Generation (B3G) should be developed mainly from the user perspective with respect to potential services and applications. Flexibility, adaptability, reusability, innovative user interfaces and attractive business models are believed according to [8], to hold the key for success of systems envisioned for deployment beyond the year 2010.

2.1.2 Service Provisioning Approaches

2.1.2.1 Intelligent Networks and IN

Mobile networks have been built upon the work of Intelligent Networks (IN) [19] from the fixed networks and has developed its own variation called Customised Application for Mobile Enhanced Logic (CAMEL) [20]. CAMEL provides a means to link the setup or management in the Mobile Switching Centre (MSC) and control the operation of the call. The GSM CAMEL feature has been developed so that operator specific services can be supported for consumers who roam to foreign GSM networks.

IN in the fixed world is mature and is responsible for delivering a range of services, but has never lived up to its promise of delivering an efficient and rapid mechanism for reusable services, as IN in general has been too difficult to implement. This difficulty is in part due to the disparity in Public Switched Telephone Network (PSTN) implementations, which has meant that while typically the majority of IN services are standard, some part always needs some form of specialist implementation to the switch or interface and this often delays service roll out and increases cost. Mobile networks are however more standardized and should not suffer from the same problems. In
mobile networks, IN distributes previously switch embedded functions across network components allowing open procurement for different components. Thus service execution and interactive voice response platforms may be bought from different switch manufacturers and from suppliers that are more computing oriented. However this has brought increased complexity due to the internetworking of different switch functions which necessitates expense non-circuit related signalling. The technology used in IN attempted to allow application developers to develop services in manner that did not require knowledge of the underlying network infrastructure but did not succeed as IN in general has always been particular to the telecommunications industry. IN/CAMEL solutions have been in part successful in providing a means to realize services in 2G and fixed networks. However, while the concepts of open control interfaces and interactions are reusable in 3G networks, service creation and control needs to be less specialized in the technologies it uses so that there is an improvement in IN [19].

2.1.2.2 Universal Mobile Telecommunication System (UMTS) and the IP Multimedia Subsystem (IMS)

UMTS [21] represents the current generation of mobile communications infrastructure and has evolved the mobile core network towards an all IP technology with a new radio network that provides higher capacity and data rates required for the support of advanced multimedia services. The new radio access network is based on Wideband-Code Division Multiple Access (W-CDMA) and has a packet-switched IP-based transport and service platform, the IP Multimedia Subsystem (IMS) [22], which is specified in UMTS releases 5 to 8. The 3\textsuperscript{rd} Generation Partnership Project (3GPP) IMS architecture consists of logical planes or layers which correspond to discrete functions. Each plane consists of IMS functional components that together provide the functions supported by the plane. The logical functions in IMS are divided into the following three planes: the transport, control and service planes, which allow network and service operators to develop and deploy services quickly. The IMS architecture is designed to enable this capability by providing an environment that is in contrast to the traditional vertically integrated silo network environment that supported individual services. The single converged network environment created by IMS eliminates multiplicity of services by enabling sharing of services across the different functional planes thereby reducing cost and creating better user experience. For further information on IMS and its functional elements see Appendix D.1.
An implementation of the service layer of the IMS has also been defined by the Open Mobile Alliance (OMA) [23] is the Service Environment specification [24], which aims to enable: interaction between different components and applications developed by different providers; specification of enablers to reduce deployment efforts and allows the same applications to operate across a wide variety of environments in a consistent manner; and specification of enablers that allows reuse so that commonly used functions can be provided for by standard components, instead of recreating those same functions in each application. It is envisioned in [24] that these enablers can be Parlay Application Programming Interfaces (APIs) such as Parlay X and Parlay applications; and an Execution Environment which deals with aspects such as Life Cycle management, load balancing and OA&M.

2.1.2.3 Application Servers and Gateway Interfaces

In UMTS Application Servers (AS) is the key method for delivering services within an operator's network. In using application servers, however, network operators need to be assured that opening up their network by defining an Application Programming Interface (API) does not expose the communications infrastructure to unauthorised use or threats. Elimination of this threat is one of the key functions of standardized middleware and gateway interfaces such as Open Service Access (OSA) [25, 26], Parlay [27], Parlay-X [28], and JAIN [29].

The Open Service Access (OSA) [25, 26] defined by 3GPP in Technical Specification (TS) 23.198, provides a standardised, extensible and scalable interface that allows for inclusion of new functionality in the network with a minimum impact in the applications using the OSA interface. Network functionality offered to applications is defined as a set of Service Capability Features (SCFs) in the OSA API, which are supported by different Service Capability Servers (SCSs). These SCFs provide access to the network capabilities on which the application developers can rely when designing new applications. Examples of SCFs offered by the SCS are call control and user location. There are three different types of OSA functions, which include Framework, network and user data functions. The framework functions provide the essential capabilities that allow OSA applications to make use of the service capabilities in the network. The framework also supports the ability for applications to access SCFs in another network. There are three distinct features that comprise the framework: Trust
and Security, Service Registration and Discovery functions and Integrity Management. The network functions represent the total collection of network resources, and include call and session control and charging functions. These are important as the application may request the quality of service when first negotiated at the start of the call and may also request the network to notify the application of any changes in QoS (conversational, background, interactive and streaming class) which take place during the call. User data related functions are subdivided into user status, user location and terminal capabilities functions. User status functions enable an application to retrieve the users’ status and notification of user status change. User location functions provide an application with details concerning the users’ location and can have a number of attributes including the location, accuracy and age of the location information. The OSA location function can also notify the application of a location update.

Using gateway interfaces such as OSA/Parlay [27] allows for rapid service creation using off-the-shelf development tools such as Borland J-Builder with Ericsson’s Open API solutions, but still requires an understanding of the underlying network infrastructure to develop new services. This requirement of knowledge cannot be assumed from an average IT software developer. In addition the adoption of CORBA as the distribution mechanism used in OSA/Parlay creates a number of problems when crossing enterprise and service provider firewalls. With the emergence of Web services into the field of telecommunications, the Parlay X Web Service API has been defined jointly by ETSI, the Parlay Group and 3GPP.

The Parlay X API [28] is fast becoming the predominant gateway interface employed by operators and aims to leverage Web service capabilities in how they are exposed as ‘services’ to the application developer. Parlay X services follow the “find, bind, invoke” paradigm by allowing services to be published and discovered through a service registry, and invoked through the discovered WSDL document. Using the Web service architecture, Parlay X simplifies how application developers can include a telecommunications call, messaging or presence features within their applications. Applications can access public network capabilities where the Parlay X service API has been published and exposed. The architectural approach of Parlay X is a means of exposing communication capabilities to a Web service application world in a way that’s familiar to that world, making it possible for applications to be built by developers.
without detailed telecommunications experience. In addition Parlay X does not use intra-enterprise integration techniques such as CORBA or Remote Method Invocation (RMI) but instead uses HTML and decouples application logic from data by using XML for representing any kind of data.

Another development is the standardization effort by the European Computer Manufacturers Association (ECMA)\(^1\) for Computer Supported Telecommunication Applications (CSTA). ECMA-323 specifies an XML protocol for services that provide XML encodings for all telecommunication service features in ECMA-269. ECMA-348 WSDL provides the Web Service specification for CSTA services. It specifies an abstraction layer for telecommunication applications, which is independent of underlying signaling protocols (i.e. SIP, H.323) and equipment. ECMA-366 specifies Web Services in WSDL and a SOAP binding for application session services allowing applications to maintain a relationship with application servers, to overcome domain based differences in stateful/stateless session based control of services, called WS-Sessions.

\(^{1}\) http://www.ecma-international.org
2.2 Marketplaces

To realise Adams Smith goal of efficient allocation of resources a market based approach to service provisioning is required where according to Feldman [16] adopting an eMarketplace approach to service provisioning will also help to improve economic efficiency, reduce margins between price and cost, and speed up complicated business deals, where the services they provide will expand many companies purchasing and selling abilities and will make processes more dynamic and responsive to economic conditions.

To enable transactions within eMarketplaces electronic Commerce (eCommerce) is employed allowing entities to conduct its business over the Internet [30]. According to the nature of such transactions, the following types of eCommerce are distinguished: Business-to-Business (B2B), Consumer-to-Consumer (C2C), Consumer-to-Business (C2B) and Business-to-Consumer (B2C). B2C refers to online retail transactions where the buyers are individual consumers and the sellers represent themselves as business cooperation's, whereas B2B refers to the transactions where both buyers and sellers are business cooperation's. Most of the initial Internet based eCommerce was focused on B2C markets, however, nowadays B2B transactions constitute a much larger portion of the overall eCommerce landscape. A Canadian survey on Electronic Commerce and Technology in 2006 demonstrated this view with 68% of all online sales being B2B related [31].

According to He et al [30] it is important to classify eMarketplaces according to the some attribute, where [30] defines the most important classification attribute to be the negotiation attribute [32]. In negotiations the topology can be classified according to:

- **Nature of interactions between agents** – which is important for an eMarketplace to distinguish whether participants are allowed to negotiate on a multilateral basis i.e. with several other participants or not. On either side – on the buyer or sellers side, one or more participants may be negotiating. Denoting the seller as M (“Merchant”) and the buyer as C (“Consumer”), Figure 2-1 shows the three possible situations given by models A, B and C.
• *Number of negotiating factors* – is an important characteristic in every negotiation as it represents the dimension of the space of negotiation issues. In more complicated “real” cases, a number of issues relating to price, quality, penalties, terms and conditions may be discussed i.e. multidimensional.

• *Whether the negotiation constraints are fuzzy or crisp* – the preferences regarding the negotiation issues may also be represented as either crisp or fuzzy, which makes it possible to evaluate a proposal and generate a counter proposal based on a certain strategy. If the issues are crisp then the preferences for these issues cannot be changed to generate a proposal or counter proposal, where if the issues are fuzzy then the various entities can truly negotiate by proposing values outside of their preferences.

Using the above attributes, Kurbel [32] developed a classification scheme presented by using a technique of morphologic boxes, as shown in Table 2-1, where the field ‘Type of Negotiation’ corresponds to nature of interactions between entities i.e. A, B or C denoted in Figure 2-1. In addition to the classification technique presented in [30], Guttmann et al [33] outlined how it is useful to explore the roles of agents as mediators in B2C and B2B eCommerce in the context of a common model, such as the Customer Buying Model (CBB) and the Business Buyer Model (BBT), which are shown in Figure 2-2 and Figure 2-3 respectively.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Possible Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of e-marketplace</td>
<td>B2B</td>
</tr>
<tr>
<td>Type of negotiation model</td>
<td>1:n (A)</td>
</tr>
<tr>
<td>Negotiation Issues</td>
<td>One issue (price)</td>
</tr>
<tr>
<td>Type of consumers constraints</td>
<td>Crisp</td>
</tr>
<tr>
<td>Type of merchants constraints</td>
<td>Crisp</td>
</tr>
</tbody>
</table>

Table 2-1 Classification of controlled multi-agent e-marketplaces

Figure 2-1 Three models of competitive negotiation in eMarketplaces
As the service provisioning platform as discussed in Chapter 1 functionality falls under negotiation model A as shown in Figure 2-1, it is necessary to classify eMarketplaces under this model. These eMarketplaces will then be further sub-classified using the classification technique outlined in [30] and [33] shown in Table 2-2 and Table 2-3 respectively.

Andersons Consulting’s BargainFinder [34] was the first shopping agent for on-line price comparisons. Given a specific product, the BargainFinder agent requests its price from nine different merchant Web sites using the same request from a Web browser. The retailers play passive roles in this process, they just provide information to the buying agents. Although a limited proof of concept, BargainFinder offers valuable insights into the issues involved in price comparisons in the on-line world. However, value added services that merchants offer in their Web sites are bypassed by
BargainFinder as it compares merchants based on price alone. Recently some online shopping markets employing this methodology were implemented. Strictly speaking they are not multi-agent eMarketplaces because the merchants are statically represented through information about their products and not through software agents. Neither are the consumer’s agents sufficiently intelligent as they possess some autonomy and very little features for cooperation. Nevertheless, some of these online shopping markets can be regarded as important steps on the way to multi-agent eMarketplaces with the negotiation model A.

Another similar example to BargainFinder is Priceline\(^1\) which carries out the same set of tasks for airline tickets, hotel rooms and cars. However a more important contribution within this domain is Jango [35], which can be viewed as an advanced BargainFinder providing a more intelligent solution by having the product requests originate from each of the consumers' Web browsers instead of from a central site as in BargainFinder. Jango’s modus operandi is simple: once a shopper has identified a specific product, Jango can simultaneously query merchant sites for its price. The results allow a consumer to compare merchant offerings based on price. However in many cases price is not the only important factor to the user. Other relevant issues, for example, might include delivery time, warranty and gift services. Also many merchants prefer their offering not be judged on price alone. Naturally the importance of different attributes will vary between consumers and so there needs to be a way for this information to be easily conveyed to the agent.

This limitation was overcome in the Frictionless\(^2\) scoring platform, “vendor scorecards” a form of multi-attribute auction that was used to measure the performance of suppliers. For example, when evaluating the performance of different laptop computer suppliers, the key factors considered include reliability, responsiveness, environmental friendliness and business efficiency. A total score is then calculated for each supplier based on the weighted score of these individual constituent components. Although quick and easy to use, the Frictionless engine neglects one essential aspect of decision making in a vague environment with fuzzy constraints and preferences. A consumer has no means to enter into the system how important the different negotiation issues or product features compared to each other. All are assumed to be equally important. This problem

---


23
was tackled by the Active Buyer's Guide System developed by Active Research, Inc. [32] The users are not only asked how desirable are certain product features for them but also how important is each product feature is when compared to others, and even how important are certain combinations of features compared to other combinations.

Two further eMarketplaces under negotiation model A are MAGMA [36] (Minnesota AGent Marketplace Architecture) and MAGNET [37] (Multi-Agent NEgotiation Testbed) developed by University of Minnesota. MAGMA was an attempt to develop a prototype of an agent-based eMarketplace together with additional infrastructure including a banking system, communication, transport and storage system, plus administrative and policing systems. MAGMA includes trader agents, which are responsible for the buying and selling of goods and negotiating prices, and an advertising server for searching and retrieving adverts by categories. Negotiation is based on the Vickrey auction, where bids are submitted in written form with no knowledge of bids from others where the winner pays the second highest amount. In contrast to the MAGMA system, the MAGNET eMarketplace was intended to provide support for complex agent interactions such as automated contracting in supply-chain management. Evaluation of the bids received is based not on cost but also on time constraints and risk, providing a very simple multi-issue negotiation technique.

MIAMI Marketplace (MIAMAP) [38] is an open virtual eMarketplace where agents process their marketing transactions, providing a generalised mediation model that supports a variety of transactions types, from simple buying and selling to complex multiagent contract negotiations. The negotiation strategy presented from this work takes advantage of the services located within the market to construct beneficial contracts. In its findings the authors in [38], state that the introduction of an explicit mediator can help resolve conflicts and add value to multiagent contracting.
Table 2-2 Classification of multi-agent e-marketplaces with negotiation model A

<table>
<thead>
<tr>
<th>CBB Model</th>
<th>Bargain Finder</th>
<th>Jango</th>
<th>Friction.</th>
<th>Active Buyers</th>
<th>MAGMA</th>
<th>MAGNET</th>
<th>MIAMAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need identification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product Brokering</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merchant Brokering</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Negotiation</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>Purchase and Delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product service and evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2-3 B2C Mediation activities of e-marketplaces with negotiation model A

Another eMarketplace with negotiation model A is the Digital Marketplace (DMP) [12], the classification of which according to [30] is shown in Table 2-4 and [33] is shown in Table 2-5, which is the only other market based telecommunication approach enabling consumers to purchase calls on a per call basis. Internally, the DMP adopts an eMarketplace where Buyers, service providers and network operators are represented by their respective agents, where agents are defined according to Woolridge[39] as "autonomous problem solvers that can act flexibly in uncertain and dynamic environments in order to achieve their design objectives", and are often required to form themselves into collectives i.e. Virtual Organisations, to act in a co-ordinated manner.
Within their embedded environment, agents can receive inputs related to the state of their environment through sensors and can act on their environment through effectors, and can be both reactive - able to respond in a timely fashion to changes that occur in their environment and proactive - able to opportunistically adopt goals and take initiative, enabling them to make context based decisions. These characteristics make them ideal in the diagnosis of problems that emerge from changes in context, as they are capable of responding to these changes efficiently and effectively.

Within the DMP the representative agents are called User Agents (UA), Service Provider Agents (SPA) and Network Operator Agents (NOA). The UA are responsible for acquiring the mobile user’s preferences over attributes such price and QoS. Upon receipt of this request the UAs initiate an auction with the SPAs using a variant of First-Price Sealed-Bid (FPSB), where the buyer selects the bidder which maximises its objective function, while meeting its valuation. Although the system allows the User Agent (UA) to specify their requirements from a multi-attribute perspective, when the Service Provider Agent (SPA) receives the request it does not formulate a bid based on these attributes. Instead it responds with a single attribute, price when is then used by the UA along with the SPA performance rating (or commitment) to determine the winner of the auction round. This limitation inherently lies in the auction protocol chosen, FPSB, where it prevents the user from correctly evaluating, what it wanted in the original request to what it actually received in the call in terms of these attributes. It also prevents the UA in performing a proper comparison between the various service providers.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Possible Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of e-marketplace</td>
<td>B2B</td>
</tr>
<tr>
<td>Type of negotiation model</td>
<td>B2C</td>
</tr>
<tr>
<td>Negotiation Issues</td>
<td>C2C</td>
</tr>
<tr>
<td>Type of consumers constraints</td>
<td>1:n (A)</td>
</tr>
<tr>
<td>Type of merchants constraints</td>
<td>M:1 (B)</td>
</tr>
<tr>
<td></td>
<td>n:m (C)</td>
</tr>
<tr>
<td>Issues</td>
<td>One issue (price)</td>
</tr>
<tr>
<td></td>
<td>Many Issues (partially)</td>
</tr>
<tr>
<td>Type of consumers constraints</td>
<td>Crisp</td>
</tr>
<tr>
<td></td>
<td>Fuzzy</td>
</tr>
<tr>
<td>Type of merchants constraints</td>
<td>Crisp</td>
</tr>
<tr>
<td></td>
<td>Fuzzy</td>
</tr>
</tbody>
</table>

Table 2-4 The Digital Marketplace (DMP) Characteristics
CBB Model | Digital Marketplace (DMP)
---|---
Need identification |  
Product Brokering | ✓
Merchant Brokering | ✓
Negotiation | ✓
Purchase and Delivery |  
Product service and evaluation |  

Table 2-5 CBB Model for DMP

In addition to the above, online auctions are no doubt the largest class of Internet-based eMarketplaces. There are literally thousands of auctions both in the B2B, B2C and C2C areas. Bean and Segev in [40] examined 100 online auctions and analyzed their characteristics. Examples of these marketplaces include eBay and Amazon, which both use a variant of the English auction to sell its goods over the Internet. In eBay to sell something one has to provide a description of the item together with some constraints including payment method, where to ship, who will pay for the shipment, minimum bid and reserve price. In fact by providing this information the seller initializes an agent to negotiate about one issue – price. For the bidder they can employ a “phantom” bidding service that utilizes the common bidding strategy of ‘sniping’ i.e. bidding at the last moment, giving the buyer the optimal chance of being “cunninger than the rest”[1]. Such examples include eSnipe and Phantom Bidder. The Fishmarket [41] electronic auction house is another example of an eMarketplace that uses the age-old institution of a fish market using the Dutch bidding protocol.

2.3 QoS based service selection

A key feature of the contribution presented in this thesis is that the various stakeholders in the system i.e. the mobile user and the service providers, can dynamically select services depending on their price and QoS constraints. The following section will outline the state of the art relating to service selection techniques from the mobile user and service providers perspective.

2.3.1 Mobile users selection of services

Mobile users within telecommunications at present have very little choice with regards to service selection based on their preferences relating to QoS or otherwise, as a result of the limitation imposed by subscription contracts.

To help users exert their bargaining power, over the past number of years various attempts have tried to resolve this problem in the mobile domain by allowing mobile users to dial a prefix before the destination number when making international calls. Companies like Swiftcall (www.swiftcall.ie), and Primus Telecommunications (www.primustel.com) enable customers to purchase international calls using a prepaid calling card or monthly bill. ZONE1511 (www.zone1511.com.sg) allows customers to choose the most appropriate supplier from a wide selection of International Direct Dialing long-distance carriers, including Cable & Wireless, China Motion, KDDI, SUNDAY and Wharf T&T. Web activated telephony such as JAJAH (www.jajah.com) provides an applet where one enters the landline/mobile number of the person to be called and the system initiates the call from the chosen endpoints. However none of these approaches allow the mobile user to specify their preferences over a host of QoS attributes.

Bearing the mobile users QoS preferences in mind in service selection procedures the most relevant works in this area is again the DMP [12] and I-centric communications [42]. In I-centric communications QoS based selection has being outlined by the Wireless World Research Forum (WWRF) as a goal [42], where the WWRF\(^1\) overall aim is to develop a common vision for future wireless systems to help drive research

---

\[^{1}\text{http://www.wireless-world-research.org}\]
and standardization. Service capabilities such as: personalization, ambient awareness and adaptability from part of I-centric communications. In adaptability services and applications are expected to change their behavior when circumstances in the execution environment changes, where adaptive applications are based on models of user preferences and models of QoS.

### 2.3.2 Service Providers selection of services based on QoS

For service providers a key requirement of service provisioning platforms for B3G was that it must enable them to become more flexible and adaptable in their business processes in order to respond rapidly to changing user requirements and competitive challenges. Service Oriented Architecture (SOA), the architecture of which is shown in Figure 2-4 has being highlighted in [43] as the technology enabler to realize this requirement.

![SOA Architecture](image)

**Figure 2-4 SOA Architecture**

Collaborations within SOAs, as shown in Figure 2-5, follow the "find, bind and invoke" paradigm, where a service consumer performs dynamic service location by querying the service registry for a service that matches its criteria. If a service exists, the registry provides the consumer with the interface contract and the endpoint address for that service. Invocation of the located service at the service provider then follows the conventional request/reply mechanism. Services with the above characteristics are referred to, collectively, as Web services and its standardization process is revolved around four organizations – World Wide Web Consortium (W3C), the Organization for the Advancement of Structured Information Standards (OASIS), Liberty Alliance and
Web Service Interoperability Organization (WS-I). WS-I is not a standardization body per se, but it combines different Web services pieces in an installation ready package, called Web service ‘profiles’ and offers tools and guidelines for installing them. The first of these profiles, called the Basic Profile defines three main standards: Simple Object Access Protocol (SOAP), Web Service Description Language (WSDL), both of which are defined by W3C and the Universal Description Discovery Integration (UDDI) standard defined by OASIS. For further information on the Web services standardization and protocols please see Appendix B.1.

SOAP is a lightweight protocol intended for exchanging structured information in a decentralized, distributed environment between the service provider and requestor. WSDL defines XML grammar for describing network services as collections of communication endpoints capable of exchanging messages, while UDDI provides a green pages directory service and adopts a standard based approach to locate services, to invoke that service, and to manage metadata relating to that service. SOAP, WSDL and UDDI standards form the basis by which arbitrary services can be defined, discovered and invoked in terms of their interfaces rather than their implementation [43].

A key aspect related to Web services is that they can contain functional and non-functional attributes and can be atomic or composite (where it is packaged as a reusable component for use in a business process) [44]. Functional attributes of a service provides a description of the service, while non-functional attributes is anything that exhibits a constraint over its functionality. Non-functional attributes are typically QoS
related. Key terms related to service composition include aggregation, orchestration and choreography.

Aggregations combine multiple services and provide access to them in a single location. Telecommunications companies can be considered as service aggregators, where services such as call forwarding, call diversion, and voicemail are brought together and offered via the telephone [44]. Service orchestration on the other hand, describes executable business processes that interact with internal and/or external services in predefined order to perform business logic. Within a composition, each sub-service is a service in its own right and complex inter-relationships may exist between the sub-services. Examples of Web and WS-Resource orchestration languages include the Pi-Calculus model with XLANG [45], IBM Petri Net model with WSFL [46] and the Business Process Management Initiative unifying the two approaches of XLANG and WSFL with BPML 1.0 [47]. The most popular and widely adopted orchestration language however is the Business Process Execution Language (BPEL) defined by OASIS [48].

Web services can also participate in a workflow, where the order in which the messages are sent and received affects the outcome of the operations performed by a service. This notion is defined as service choreography. The difference between orchestration and choreography is shown in Figure 2-6. Orchestration always represents control from one party’s perspective and differs from service choreography, which is more collaborative and allows each involved party to describe its part in the interaction. Choreography tracks the message sequences between multiple parties and sources – typically the public message exchanges that occur between the Web services - rather than the specific business process that a single party executes [49]. W3C specification, the Web Service Choreography Description Language (WS-CDL) [50] is a choreography language that “defines from a global viewpoint ... the information exchanges that occur and the jointly agreed ordering rules that need to be satisfied”.

31
A significant characteristic of a service composition strategy is the degree of automation in the creation of a process model. Traditional service composition methods require the user to define the data and control flow of a composite service manually, either directly or by means of designer tools. Most of these techniques fall under the domain of workflow composition.

Authors point out in [51, 52], that workflow systems do not cater for dynamic and distributed nature of service composition for two main reasons: (1) a common workflow modelling and management environment is impossible across different enterprises since no workflow management systems share the same workflow syntax and semantics; (2) workflow systems do not offer facilitates such as changing flow definitions which is a fundamental requirement for service composition. Therefore workflow solutions may work for semi-fixed or fixed compositions, however they do not work well with explorative composition which requires the service composition structure to be generated on the fly and the composition itself to be changeable [53].

Fully-automated composition approaches intend to generate a service composition plan without human intervention. Within AI planning several methods for service composition have been proposed using techniques from: situation calculus [54] and rule-based planning [55]. Works from the domain of situation calculus include [56, 57]. In [57] the authors argue that an augmented version of the logic programming language Golog [58] provides a natural formalism of automatically composing services on the Semantic Web. These contributions are realized in their development of the ConGolog [59] interpreter which communicates with Web services via the Open Agent
Architecture (OAA) but the service and procedure ontologies are written in first-order logic. Within Planning Domain Definition Language (PDDL) provided by Ghallab et al [60], a strong interest to Web service composition from the AI planning community can be explained by similarities that exist between OWL-S and PDDL representations. PDDL is widely recognized as a standardized input for state-of-the-art planners. When planning for service composition is needed, OWL-S descriptions could be translated to PDDL format, and then different planners could be exploited for further service synthesis. Such works include those produced by Klusch et al [61, 62] who uses a hybrid based approach AI planner Xplan and Wu et al [63] SHOP2 planner. Within Rule-based planning Medjahed [55] presents a technique to generate composite services from high level declarative description. This method uses composability rules to determine whether two services are composable. The composition approach consists of four phases. In the first phase, the specification phase enables high-level description of the desired compositions using a language called Composite Service Specification Language (CSSL). Secondly, the matchmaking phase uses composability rules to generate composition plans that conform to service requester’s specifications. If more than one plan is generated, in the selection phase the service requester selects a plan based on quality of Composition parameters. The final phase is the generation phase, where a detailed description of the composite service is automatically generated and presented to the service requester. The main contribution of this method lies in the area of composability rules, because they define the possible Web service’s attributes that could be used in service composition.

Within the area of dynamic service composition for Web services there are a number of issues that have arisen within the domain due to the dynamic nature of Web services. These dynamics include: market and service capability feature flux, where the set of services capable of providing functionality can change constantly; selection of the most appropriate business process where more than one plan can exist; changing service properties where a business process needs to accommodate itself to system irregularities such as failure; and the performance and execution of a business process which is measured in terms of its overall end-to-end Quality of Service (QoS). For these reasons service provisioning platforms in the future needs to make service composition and consequently service provider collaborations decisions using a well defined system model where it can:
1. Appropriately determine the service providers preferences over a range of QoS attributes such as performance, rating and availability

2. Use the preferences as determined from the first stage to identify the most capable and effective atomic service to participate in the business process

3. Compose and bind the model using semantics and ontologies.

The scope of this thesis however deals specifically with the first two parts of the system model and does not deal with the AI aspects relating to the generation of the corresponding process and execution model using semantics and ontologies.

Because of the large number of automated service composition approaches as outlined above are available, within the contribution of this thesis a general framework for automatic Web service composition proposed by Rao & Su [64, 65] as shown in Figure 2-7. A generic composition system has two types of participants, service provider and service requester. The service providers propose the Web service for use. The service requesters consume information or services offered by service providers. The system also contains the following components: translator, process generator, evaluator, execution engine and service repository. The translator translates between the external languages used by the process generator. For each request, the process generator tries to generate a plan (or workflow graph) that composes the available services in the service repository to fulfill the request. In the presence of multiple service providers with comparable or identical functionality, users will want to discriminate these alternatives based on their desired QoS. According to Jaeger et al [66] and Fung [67] this selection should be made on the non functional attributes related to the service. Quality has been regarded as a driver of competitive strategy, and for strategic purposes, many quality frameworks have been developed. For instance Garvin [68] developed an eight dimension quality framework, while Parasuraman et al [69] specified additional service quality dimensions. Within this problem area there have been a number of contributions, the most relevant of which are described below.
Zeng et al. [70] presents a middleware platform which addresses the issue of Web services selection for the purpose of their composition that maximises user satisfaction expressed as utility functions over QoS attributes. The approach uses task level selection and integer programming techniques. However the approach presented is too time consuming for real time scenarios in eBusiness, which was confirmed by work conducted by Canfora et al. [71] and Yu and Lin [72]. Yu and Lin propose a QoS capable Web service architecture where the problem is modelled as a multi-choice knapsack problem. A similar approach in [73] describes FUSION for dynamic Web service composition and automatic execution. The project’s goal is to automatically generate an optimal execution plan from user’s requirements, modelling the problem again as a Multiple Choice Knapsack Problem. Canfora discusses the use of genetic algorithms as an approach for solving Web service composition.

Wang et al. [74] presents a multidimensional model called MDF4SS, that evaluates the properties of Web services given a set of user requirements and a set of candidate services and adapts to changes during the service selection by revising the dimensional attributes in order to conform with the user constraints. The work presented to date however has not presented a mathematical model how to achieve this, and while it does aim to provide a ranking service for Web service selection it does not specify the use of e-marketplaces, agents and adaptive strategies.

CIM [75] and eFlow [76] discuss the possibility of performing dynamic service selection based on user requirements. CIM’s service definition model features the concept of a placeholder activity to cater for dynamic composition of services. In
eFlow, each service node contains a search recipe, which defines the service selection rule to select a specific service for the node. CIM and eFlow focus on optimising service selection at a task level. In addition no QoS model is explicitly supported. In contrast, the TSE provides a well defined model for QoS support through the use of auctions and preference elicitation techniques.

Other related work on QoS based composition has been done in the area of workflow. Patel et al [77] propose a QoS oriented framework, WebQ, for adaptive management of Web service based workflows. Several QoS parameters such as latency, throughput, reliability, availability and cost have been defined and a selection is made based on rule based systems. Most efforts in this area focus on specifying and enforcing temporal constraints [78]. Some projects such as METEOR [79] and CrossFlow [80] consider other QoS criteria. In the METEOR project four QoS attributes are defined: time, cost, reliability and fidelity. However, it does not consider the dynamic composition of services, it instead focuses on analysing, predicting and monitoring workflow QoS. CrossFlow proposes the use of continuous Markov chains to estimate execution time and cost of workflow instances.

Priest [81] proposes the use of auctions for service composition however it suggests the use of the English auction protocol and does not consider QoS attributes in the selection. Esmaeilsabzali et al [82] adopts a game theoretic approach to Web service selection in designing a multi-attribute mechanism where two quality attributes, namely reliability and availability are considered.
2.4 Summary and Motivation for the TSE

Based on the original strategic requirements of the regulating authorities and the state of the art presented in this chapter, it is clear that there is no service provisioning platform in existence that can meet the regulators objective of effective competition by enabling: service providers to quickly develop services that the mobile users want and; mobile users to switch easily between suppliers.

Current approaches to service provisioning such as the IP Multimedia Subsystem (IMS) lets carriers roll out multiple applications over one network without having to build a separate network for each application. IMS gives carriers the potential to realize enormous savings because they can reduce the amount of equipment in the network and reduce the number of people needed to support it and also provides carriers with the flexibility to respond more quickly to new business opportunities and offer revenue-generating applications. In addition, standardization efforts by ETSI, the Parlay Group, 3GPP and ECMA also represents a new direction for network operators, facilitating the use of telecommunication services by Internet developers and provides a tighter integration of telecom services and the Web.

However, while the IMS and gateway interfaces offers significant opportunities for network operators it does not utilize some of the key enablers for dynamic service composition such as agents, electronic Marketplaces, semantic technologies and automated negotiation. Such technologies are required for the full automation of the broad spectrum of activities related to dynamic service composition making service providers fully responsive to ever changing user requirements and competition and would enable the true potential of the IMS and its functional service planes. In addition the IMS or gateway interfaces does not allow service and network operators to form collaborations with other service providers dynamically through interoperable software components over the Internet and be charged for such a service on an as needed basis, nor does it enable mobile users to exert their bargaining power by allowing them to purchases services on a per request basis.

Within eMarketplaces approaches such as BargainFinder [34], JANGO [35], MAGMA[36], MAGNET[37] and MIAMAP[38] provided useful insight into the development issues relating to eMarketplaces, however they lack a specific context to
the telecommunications in addressing the problems of service provisioning and mobile
users inability to exert their bargaining power. The DMP [12] eMarketplace on the other
hand, is specific to the telecommunications domain and is the only other market based
approach enabling mobile consumers to purchase calls on a per call basis.

The DMP uses a Logical Market Channel (LMC) [12], which operates as a random
access channel on the MAC layer. Users gain access to the DMP via the LMC, using
radio resource registration, connection and termination messages. The use of the LMC
requires modification of the air interface, which brings with it a host of new regulatory
and standardisation challenges. The TSE on the other hand uses the Session Initiation
Protocol (SIP) [83] in all its registration, connection, termination and transfer
procedures.

The advantage of using SIP is that it is a session based, access-independent application-
layer protocol, compared to the MAC layer approach with the LMC and radio resource
signalling in the DMP. SIP is session-based rather than call based giving network and
service providers the freedom to offer innovative services that let subscribers add users
and media at will. A session could be an IP telephony call, a multi-user conference that
incorporates voice, video and data, instant messaging chat or a multi player online
game. By using the key features of SIP, service providers can focus on developing the
aspects of a service that will most enrich the user's experience. The details of the
session, such as the media, codec or sampling rate are not defined using SIP. Rather the
body of the message contains a description of the session, encoded in some other
protocol format, such as the Session Description Protocol (SDP) [84]. Its extensibility
and wide use also in Voice over IP (VoIP) applications made it the natural choice of
signaling protocol for the TSE. In addition, 3GPP has chosen SIP as the signaling
protocol for the IP Multimedia Subsystem (IMS) of Universal Mobile
Telecommunication Service (UMTS) [21] networks, and hence was the natural choice
for signalling in the TSE. Further information on SIP is provided in Appendix A.

In negotiating, the DMP using a variant of First-Price Sealed-Bid (FPSB), where the
buyer selects the bidder which maximises its objective function, while meeting its
valuation. As stated in Section 2.3 the DMP system allows the UA to specify their
requirements from a multi-attribute perspective but formulates a bid/response based on
price as the unique strategic dimension. This limitation inherently lies in the auction
protocol chosen, FPSB. It is the aim of the TSE to enable negotiation over a range of attributes and where negotiation and proposals made can be formulated based on fuzzy constraints, enabling true negotiation.

The DMP also does not support dynamic service composition procedures enabling service providers to dynamically collaborate with each other based on their price and QoS requirements. It is the aim of the TSE to specifically address this problem, where coalitions between service providers can be established and terminated on the fly depending on the performance of their corresponding atomic service elements.

Other approaches which aim to enable the user purchase a call on a per call basis outside on their subscription contracts include prefix dialing and Web activated telephony which are services provided by companies such as Primus Telecommunication, ZONE1511 and JAHAH. While the above approaches do give mobile users some choice in local and international markets for making mobile calls, it may not give the mobile user the best deal possible. ZONE1511 and JAHAH tariffs are openly displayed and demonstrate little competitiveness within the sector. In addition such approaches are again call based and not session based, which is a serious disadvantage as mobile users in the future want much more than basic telecommunication services such as mobile calls as previously outlined in Chapter 1.

The transactions relating to dynamic service composition within the TSE relate to plan or component selection for participation in a business process. As a result of this a state of the art conducted focused on QoS based selection of Web services. Work from Zeng et al [70], Wang et al [74], CIM [75], eFlow [76], Patel et al [77], METEOR [79] and Crossflow [80] attempt to resolve this problem by using techniques such as task level selection, integer programming techniques, adaptive strategies and representing the problem as a Multiple Choice Knapsack Problem. Priest [81] proposes the use of English auction protocols but the choice of protocol prevents the mobile user or service provider specifying multiple attributes that form part of their decision process, as only price is considered to be important. The closest work as that presented in this thesis was Esmaeilsabzali [82] who proposes to use a multi-attribute auction protocol. However the approach undertaken focuses primarily on service provider strategies, but does not evaluate how the service providers' weights are determined as inputted from the fuzzy constraints by the user which is a key requirement in any negotiation model. In addition,
the work presented in [83] does not provide a framework where the service providers and the transactions relating to Web service composition can be supported.

Based on the above it is clear that there currently does not exist a service provisioning platform that incorporates the key features required for such as outlined in Chapter 1. As a result it is the objective of this thesis to present the Telecommunication Service Exchange (TSE) to overcome the limitations of the current approaches to service provisioning for B3G.
Chapter 3

TSE Design

This chapter presents the main contribution of this thesis, dealing with the design and description of the Telecommunication Service Exchange (TSE) framework. The chapter will commence with a description of the properties that the TSE must incorporate to realize its potential, which follows from the requirements of service provisioning platform for B3G as specified in Chapter 1 and the state of the art outlining the limitation of current approaches. Following this the various design elements will be presented and a rationale for their selection. Once the various constituent components are selected the TSE generalized and internal market architecture will be outlined.
3.1 TSE Features

After analyzing the current approaches to service provisioning, as detailed in the previous chapter, it was necessary to specify the principle requirements that the TSE must satisfy in order to enable its full potential. These characteristics are further outlined below:

1. The TSE must be *responsive* and *adaptable* in the sense that service providers can respond to changing:

   a. mobile user service requirements

   b. market and service capability flux where the set of services providing functionality can change constantly

   c. service properties where a process needs to accommodate itself to system irregularities such as failure

2. The entities within the TSE must be *autonomous* and *proactive* in the sense that they can exist independently from its owners with minimum support and are able to make their own decisions and opportunistically adopt goals and take initiative when required. The entities also must be *social* where it will be required for them to communicate with each other in order meet their design objectives

3. The TSE must provide support where mobile users and service providers *preferences* over a host of QoS and price issues are obtained in a manner appropriate to the capabilities of the various stakeholders, and where *automated negotiation* over these issues is commenced upon receipt of the stakeholders request

4. Finally the TSE must provide the *architectural, supporting* and *signalling services* to allow it to become a service provisioning platform for B3G

In order to realize the above aforementioned features of the TSE, a number of design decisions have to be made. It is the purpose of this chapter to specify these decisions along with a well reasonable rationale for its choice.
3.2 Service Oriented Architectures (SOA)

As stated in Chapter 1, SOA in telecommunications gives service and network operators the ability to view their underlying network infrastructure more as a commodity and allows infrastructure development to become more consistent. Adopting this process-centric approach to service provisioning enables faster time to market, as new initiatives can reuse existing services and components thus reducing design, development, testing and deployment time in addition to the cost and risk of undertaking such projects. Looking at the TSE required features, as specified in Section 3.1, SOA essentially enables service and network operators to become more flexible and adaptable in their business processes enabling them to respond rapidly to changing mobile user service requirements.

Research studies undertaken by Forrester [85] and IBM [86] support the above claim. IBM [86] undertook a study of 35 SOA projects, across a range of industries and regions. They discovered that every company in their study showed increased flexibility, 97% decreased their cost, 71% reported reduced risk of undertaking projects, 51% increased their revenue and 43% of companies enabled new products.

One of these companies was Sprint, who developed a locator application in response to recent United States emergency call regulations that makes cell phones location information available to emergency personnel, who can use the information to track people who need emergency assistance. Realizing the potential benefit of the service to other domains Sprint wanted to make the locator application available to other broader range of its subscribers, such as consumers for truck fleet management. To do this Sprint extended the locator application and integrated it with General Positioning System (GPS) technology, Presence, messaging and voice services, enabled through the successful adoption of a SOA approach. Successful implementation of SOA and the resulting new service, reduced development time and effort by 40-50% compared with traditional GPS application development. In addition the time to acquire a location from a cell phone reduced from potentially 6 minutes to 30 seconds, and the location information obtained was more accurate.

The Forrester [85] research report outlined how nearly 70% of SOA users say they will increase their use of SOA, and 46% of large enterprise users of SOA use it for strategic
business transformation. Between these two reports one message became clear that companies adopting SOA implicitly understand that SOA entails massive business benefits, not least in the critical area of innovation, but also in reducing development time, risk and effort in undertaking such projects.

In addition to the above, key industry players are now focusing on providing accessible SOA/Web services environments. PayPal has developed a set of APIs [87] that merchants and other users can use to automate certain functions that otherwise require manual intervention. Examples of functionality that are exposed via Web services are refund-processing, queries against transactions and secure withdrawal of funds from buyer's accounts. By making use of these services, merchants and retail users can integrate their tools to extend and tailor PayPal's functionality to meet their business needs. Google has created a set of APIs [88] that allows users to integrate Google's capabilities such as search and spell check into their systems. Using these APIs, developers can write programs that integrate with Google services using SOAP. Amazon.com has also provided a set of Web services [89] that developers can extend to provide inventory-management tools, pricing configuration tools and store building tools.
3.3 Stateful Web services

In addition to the requirement where the TSE must become more flexible and adaptable in their business processes, the TSE must also become more flexible and adaptable to changing market and service capabilities fluxes and system irregularities such as failure. In order to facilitate this feature the B2B market of the TSE needs to manage the Web services effectively.

Two working groups dedicated to the management of Web services are the Web Service Management Working Group [90] and the Web services Distributed Management (WSDM) Technical committee (TC) [91]. The Web service Management Working Group is focused on the management capabilities of Web services which include the: identification, status, configuration, metrics, operations, and events, relating to Web services. A key aspect determined from this working group was the management of a Web services lifecycle [90]. Oasis also set up a Web Services Distributed Management (WSDM) Technical committee (TC) [91] where it has published two specifications: Management Using Web Services (MUWS) and Management Of Web Services (MOWS). MUWS enables management of distributed IT resources using Web services, while MOWS focuses on the management of the Web service endpoints using Web service protocols using the concepts and definitions specified in MUWS. The WSDM specification also outlined key manageability capabilities to include: identity, manageability characteristics, correlatable properties, description, state, operational status, metrics, configuration, relationships and advertisement.

Key points in both working groups include the notions of state, event and lifecycle. However Web services by their nature are stateless and non transient. This means that the service does not keep state from one invocation to another, and does not have the concept of service creation and destruction, and hence managing its lifecycle. However the use of state, transient or persistent, is critical to the management operations of many Web services, and fundamental in building distributed telecommunication applications. In order to address this issue, services adopted in next generation service platforms should be focused on stateful and persistent Web services, called Web Service-Resources (WS-Resources), a concept that was originally developed by the Grid
community. Appendix B.2 outlines further information on standardization efforts and bodies for WS-Resources and also provides an overview on the Semantic Grid efforts.

A key feature of WS-Resources is that they are persistent, where a request when sent creates a transient process called grid service instance with a unique identifier called a handle, which can be used to locate and query the instance. The service instance exists only for a limited amount of time, after which the instance will be destroyed. If required, the client/application can also extend the expiration time of the service instance. Another key feature of WS-Resources is that they are stateful, which means they maintain information across multiple operations issued over time, providing a standard mechanism to expose the data associated with each service instance for query, update and change notifications.

To support WS-Resources statefulness, OASIS has defined the WS-Notification (WSN) [92] specification which enables the use of publish and subscribe mechanisms. WSN allows the simulation of event-driven applications in the stateless environment of Web services, and enables the Service Providers Agents (SPAs) in the TSE to know when a property of a resource has changed. WSN defines several basic roles including the Notification Producer and Notification Consumer. In the simplest of cases, the Notification Consumer contacts a Notification Producer for a Subscription to a particular topic. When the Notification Producer has a message related to that topic, it sends a message to the Notification Consumer. In a more complicated scenario, the NotificationConsumer does not have to subscribe directly. There may be a Subscriber object that creates the Subscription for it, or there may be a Publisher that can't handle the tasks of the NotificationProducer. In that case, it sends it message to a NotificationBroker, which in turn sends them to NotificationConsumer. This configuration is part of the WS-BrokeredNotification specification. The TSE uses WSN to correctly monitor the atomic and composite service execution with regards to SLA monitoring, service failure and achieved QoS characteristics. In order to deal with such notification events an Event Driven Architecture (EDA) must be adopted in the network providers' home network or internally within the TSE. Appendix B.1.3 provides further details on EDA.

Following the above discussion, in order for service and network providers to provide the management activities relating to Web services the use of WS-Resources needs to
be a stipulation stated in the TSE. In telecommunications such a distributed infrastructure provided by WS-Resources will increase the range of potential service providers and their respective services to collaborate with, reducing the development time of the niche services and increasing the service providers’ flexibility in adapting to change and meeting users’ demands. These collaborations could facilitate service providers to test such niche services and its potential demand even more quickly, without committing its own network resources, services or development efforts in building them and pay for them on an as needed basis. In essence distributed collaborations bring more choice and flexibility to the service providers in its SOA framework. In addition, telecom service and network providers can also realize significant cost savings by outsourcing nonessential elements of their IT environment to various other forms of service providers making better utilization out of their existing assets. Through the adoption of WS-Resources, service providers could also coordinate with global partners and multinational companies i.e. Vodafone Ireland and Vodafone China. Such a partnership could realize several benefits as the time zones between the partners could facilitate load balancing of service loads without having adverse affects on either service provider’s network (i.e. peak time of calls and services for Vodafone Ireland is off peak time in Vodafone China). [9, pg. 81]

According to Gartner many businesses will be completely transformed over the next decade by using Grid enabled Web services that share applications and computing capabilities. Another research report from Clabby Analytics [93] stated that “Grid and Web Services standards have advanced to the point where it is now possible to build a SOA/Grid capable of supporting dynamic business process flow”. Within telecommunications, WS-Resources are already transforming business processes by enabling service providers to compete more effectively, which are further outlined below.

A recent collaboration between Fujitsu and France Telecom, presented at the 17th Global Grid Forum, developed a system based on grid computing that enables telecommunications carriers to optimize their use of IT resources. Tests demonstrated that service loads were automatically allocated among servers in Paris, Tokyo and Kawasaki, enabling the overall system to handle loads that would have been beyond the capacity of conventional systems. Another example of WS-Resources successful
adoption in telecommunications is Embarq [94] who in its communications infrastructure in the U.S. Midwest ran a project to improve consumer relationship management and increase employee efficiency and productivity. Embarqs’ problem emerged due to the fact that most of its data sources were based on geographically distributed heterogeneous systems. Adopting grid middleware provided a virtual consolidation of ten data systems into a unified real time view of consumer information for sales and consumer service representatives.

Based on above discussion relating to the research activities of the management of Web services and the industrial success in realizing WS-Resources potential, it is envisioned within the context of the TSE that service providers and network operators will realize also its potential through the adoption of an Event Driven Architecture (EDA) and WS-Resources.
3.4 Agents

The TSE proposes the use of an open and dynamic system to service provisioning for B3G. However, with this requirement heterogeneous systems must interact, span organizational boundaries and operate effectively with rapidly changing circumstances. As a result the TSE needs a degree of autonomy to enable entities to respond dynamically to changing circumstances while trying to achieve their objectives. The TSE also needs to be flexible in the sense that entities operating within it are responsive, proactive and social in their environment. The entities also need to be proactive in their decision making where they are able to exhibit opportunistic, goal directed behavior and take initiative where appropriate.

Bearing all these design requirements in mind, service providers and network operators operating within the TSE are represented using software agents, where Woolridge et al [39, 95] defined an Agent as a “computer system, situated in some environment that is capable of flexible autonomous actions in order to meet its design objectives”. Appendix C.1 provides more information on agents their communication protocols and the theories of agencies that govern their behavior. From the perspective of the TSE the inherent characteristics of agent can help with the management of distributed resources over the Internet, Web service composition and automated negotiation which are further discussed below.

3.4.1 Applications of Agents

Within Grid computing agent based interactions have very similar characteristics to that of Virtual Organizations in the Grid, where “coordinated problem solving in dynamic, multi-institutional Virtual Organizations” is one of their main objectives. The Grid and agent communities are both pursuing the development of such open distributed systems, albeit from different perspectives. The Grid community has historically focused on what Foster et al [96] refers to as the “brawn” i.e. an interoperable infrastructure for secure and reliable resource sharing within dynamic and geographically distributed Virtual Organization (VO), while the agent community has focused on the “brains” i.e. on the development of concepts, methodologies, and algorithms for autonomous problem solvers. According to Foster et al [96], integrating the ‘brawns’ of the grid, with the ‘brains’ of the agent could result in “a framework for constructing large scale, agile
distributid systems that are qualitatively and quantitatively superior to the best practice today". The TSE aims to use this concept where the service providers and their services can be distributed across the Internet, utilizing stateful Web services and the WSN framework to provide the brawn, while the agents provide the brain to deal and respond to changes in the distributed system.

Because of the horizontal nature of agent technology, it is also envisioned according to [97] that the successful adoption of agent technology with Web services will have an profound, long term impact both on the competitiveness and viability of IT industries and also on the way in which future systems will be conceptualized and implemented. With Web services, W3C have described agents as the “running programs that drive Web services – both to implement them and to access them as computational resources that act on behalf of a person or organisation”, where works in [98,99,100,101], propose and demonstrate the successful adoption of agents for dynamic Web service composition. Hence agents provide the natural software computing paradigm for the automation procedures relating to Web services.

A key aspect within the TSE is the eCommerce and negotiation activities of the markets. Within these markets agents are also used to fully realise the economic benefits to its implementation, where according to Jennings [30] “Electronic Commerce is the most important allocation for Agent technologies, because it is reality-based and constitutes a massive market”. As a result the adoption of agent technology is central element to the operations within the TSE, where these agents negotiate on behalf of their owners.

3.4.2 Agents, Semantics and Ontologies

In order to support meaningful communication between the agents and services developed by different vendors in the TSE, a common understanding of used terms and conditions need to be achieved. But as there is no global common understanding and use of terms, this understanding needs to be establishment between parties on the fly. In order to enable this mechanism, semantic descriptions using taxonomies and ontologies need to be used.

In the context of knowledge sharing, Gruber [102] defines the term ontology to mean a specification of a conceptualization. To be able to discuss with one another,
communicating parties need to share a common terminology and meaning of the terms used. Otherwise, profitable communication is unfeasible because of the lack of shared understanding. With software systems, this is especially true – two applications cannot interact with each other without common understanding of terms used in the communication. Until now, common understanding has been achieved awkwardly by hard-coding this information into applications. Ontologies describe the concepts and their relationships, with different levels of formality in a domain of discourse. An ontology is more than just a taxonomy (classification of terms) since it can describe more than relationships between the defined terms.

Within agents instances of ontologies classes can be passed in Agent Communication Language (ACL) messages in accordance with FIPA specifications using either FIPA-SL [103], Resource Description Framework (RDF) [104,105], FIPA Knowledge Interchange Format (KIF) [106] or FIPA Constraint Choice Language (CCL) [107]. JADE provides support for the SL, RDF and XML codec’s for encoding the content of ACL Messages. To aid ontology development for agents, ontology editors, such as Protégé [108], help human knowledge engineers develop and maintain ontologies. They support the definition and modification of concepts, slots, axioms, and constraints to enable the inspection, browsing and codifying of the resulting ontology.

Within the TSE implementation RDF was chosen in preference to the more conventional use of FIPA-SL in the content of FIPA messages for a number of reasons. Firstly, choosing RDF lends greater interoperability by aligning with W3C recommendations to the semantic Web while still being FIPA compliant. Secondly, the TSE can reuse existing schemas or ontologies in a particular application domain. Thirdly, particularly at the lower (RDF) layers of the semantic Web formalism stack, the semantics of the data model are much simpler than FIPA-SL, while still adequate for operational use. In addition at present RDF is sufficiently expressive to capture usable structures, and has facilitated the rapid development and implementation of the TSE and the necessary message formats for inter agent communication.
3.5 Automated Negotiation

A key element to the TSE and its operation is the automated negotiation procedures involved in both markets, which are detailed further in the forthcoming section. A concept relating to automated negotiation is negotiation analysis, which includes key concepts such as game theory and mechanism design, which is further outlined in Appendix C.2.

Within negotiation, current human-based procedures are relatively slow, does not always uncover the best solution, and are, furthermore, according to Beam and Segev [138] constrained by issues of culture, ego and pride. Experiments and field studies demonstrate that even in simple negotiations people often reach suboptimal agreements, thereby “leaving money on the table”. The end result is that the negotiators are often not able to reach agreements that would make each party better off.

The fact that negotiators fail to find better agreements highlights that negotiation is a search process. What makes negotiation different from the usual optimisation search, is that each side has private information, and neither typically knows the other’s utility function. Furthermore both sides often have an incentive to misrepresent their preferences. Finding an optimal agreement in this environment is extremely challenging. Both sides are in competition but must jointly search for possible agreements. Although researchers in economics, game theory and behavioural sciences have investigated negotiation processes for a long time, a solid and comprehensive framework is still lacking. A basic principle of microeconomics and negotiation sciences is that there is not a single “best” protocol for all possible negotiation situations. Wurman et al [110] asserts that different negotiation protocols are appropriate in different situations, and, thus, any generic mediation service should support a range of options.

Recent developments in electronic market research offer the promise that new negotiation protocols will not only leave less money on the table but will also enable new types of transactions to be negotiated more cost effectively. There have been many approaches for supporting or automating commercial negotiations, such as bargaining and auctions. Bargaining situations can concern as few as two individuals, who may try to reach an agreement on a range of transactions. Over the past decade, there have been
several approaches to supporting or describing one-on-one negotiations, ranging from
game theory to negotiation support systems to intelligent agents who bargain the details
and finally close the deal without any further user interactions. According to Bichler
[111] however, although much research has been accomplished, automated bargaining
using agents is currently restricted to a small number of applications in commercial
environments. The reasons for this is that game theory has failed, thus far, to describe
human bargaining, where Linhart et al [112] state that “inadequate theories of
bargaining exist only for the degenerate, polar cases of competition and monopoly”. In
addition negotiation support systems require constant human input, both at the initial
problem setup and all final decisions are left to the human negotiators, making
automated bargaining not so automated.

McAfee [113] defined an auction as “a market institution with an explicit set of rules
determining resource allocation and prices based on bids from the market participants”. Auctions constitute one type of dynamic pricing, in which the price of the
product varies, depending on the demand characteristics of the consumer and the supply
situation of the seller and are often used rather than posting a fixed price on an item, in
cases where products have no standard value. In addition to price determination,
according to Kagel [114], auction theory is also important for practical, empirical and
theoretical reasons. The roots of electronic auction and negotiation mechanisms are in
auction and negotiation theory. See for instance, Raiffa [115], Milgrom [116], Kagel
and Roth [117], Klemperer [118] and Rothkopf and Harstad [119]. There are many
different forms of auctions, where [120] defines a taxonomy of auction parameters that
allows for approximately 25 million types of auctions. Beam & Segev [40] also
examined 100 online auctions and analyzed their characteristics.

However despite this vast range of auction protocols there are only four common types
of single sided auctions, which include: English, First Price Sealed Bid (FPSB), Second
Price Sealed Bid (SPSB) and Vickrey. The most common type of double-sided auction
is the Continuous Double Auction (CDA), which allows buyers and sellers to
continuously update their bids at any time in the trading period. Additional detail on
these auction protocols is outlined below:

1. **Ascending Bid or English auction**: the price is successively raised until at least one
   bidder remains. This can be done by having an auctioneer announce prices, or by
having bids submitted electronically with the current best bid posted. The essential feature of the English auction is that, at any point in time, each bidder knows the current best bid. Antiques, art work and houses are sometimes sold using this type of auction [118]. The agent’s dominant strategy (the best thing to do, irrespective of what the others do [121]) is to bid a small amount more than the current highest bid and stop when the user’s valuation is reached. For example, in Yahoo auctions, “autonomic bidding” allows users to input their maximum bid and an agent will bid incrementally when it is necessary to win the auction.

2. **Descending Bid or Dutch auction:** is the converse of the English auction. The auctioneer calls an initial high price and then lowers the price until one bidder accepts the current price. The Dutch auction is used for example, for selling cut flowers in the Netherlands, fish in Israel and tobacco in Canada. An analysis of strategies in Dutch auctions can be found in [122].

3. **First-Price, Sealed-Bid (FPSB) auction:** each bidder independently submits a single bid, without knowledge of what bids are submitted by other participants. The object is sold to the bidder who makes the highest bid. This type of auction is used in auctioning mineral rights in government-owned land, and is sometimes used is the sales of artwork and real estate. Of greater quantitative significance is the use of sealed bid tendering for government procurement contracts - that is competing contractors submit prices and the lowest bidder wins and receives her price for fulfilling the contract. The dominant strategy in FPSB of complete information is to bid the second highest bidders valuation, while in FPSB of incomplete information the dominant strategy, computed using game theory is that he bids a fraction \( \frac{n-1}{n} \) of his valuation \( v \), when a total of \( n \) parties are bidding. Further analysis of this strategy is provided in [113].

4. **The Vickrey or Second-Price, Sealed-Bid auction:** operates in the same manner as FPSB and while the object is still sold to the bidder who makes the highest bid, the winning bidder pays the second-highest bidders bid, or “second price”. While this auction has useful theoretical properties, it is seldom used in practice due to its vulnerability to a lying auctioneer, lower revenue when compared to the English
auction and undesirable private information problems [123]. The (weakly) dominant strategy used in Vickrey auctions is to bid the valuation \( v_i \) for player \( i \).

5. **Continuous Double Auction (CDA):** This type of auction is easy to operate, efficient and can quickly respond to changing market conditions. A variety of CDA models have being constructed [124, 125] and these vary in terms of whether bids/asks are for multiple or single units, whether unaccepted offers are queued or replaced by better offers and so on. Nevertheless all these protocols allow traders to make offers to buy or sell and to accept other trader’s offers at any moment during a trading period. The messages exchanged generally consist of bids (offer to buy) and asks (offers to sell) for single units of the commodity, and acceptances of the current best bid or ask. Several bidding strategies have been proposed in the literature. The ZERO Intelligence strategy [126], generates a random bid within the allowed price range decided by the agent’s budget constraint. The adaptive agent bidding strategy is based on stochastic modeling of the auction process using a Markov chain [127]. A sequential bidding agent method using dynamic programming is proposed in [128]. In [129], heuristic fuzzy rules and fuzzy reasoning mechanisms are used to determine the best bid given the state of the marketplace.

Auctions have been described by Binmore and Vulkan [130] as “an effective way of resolving the one to many bargaining problem”. According to experiments described in Kagel et al [117], the outcome of market competition is more likely to conform to game-theoretical rationality than the outcome of a bilateral negotiation. Economists view competition as a means of setting prices right, thereby generating an efficient allocation of resources. Being a bid taker puts less of a burden on the seller’s knowledge and abilities than being a negotiator in a bargaining process, simply because she does not need to know the range of possible buyer valuations. The winning bid is on average below the item’s true but unknown value, but with the introduction of more and more bidders, the price approaches its true value. The technical infrastructure required to support auctions in an online environment is currently available and well accepted. Wurman [131] outlined how auctions are a very efficient and effective method of allocating goods/service, in dynamic situations to the entities that value them most highly, whereas Bapna et al [132] stated that “Online auctions, bought about by the synergetic combination of Internet technology and traditional auction mechanisms
present a significant new dimension for mercantile processes”. Agents can also represent their owners in an auction, where the agent activities may involve monitoring, analyzing the market conditions and/or deciding when and how much to bid for the desired items. Automating these activities through the use of agents can save time, and in complex settings it has been shown by research by Das et al [133] that when agents and humans participate simultaneously in a realistic auction, the software agents consistently produce greater gains compared to their human counterparts. As a result of the above, the TSE uses auctions as the means of achieving automated negotiation procedures. Appendix C.2.4 provides further details on auctions its types and extensions.

3.5.1 Multi-Attribute Auctions

To date, most of the research on automated auctions has focused where price is the unique strategic dimension [134, 118]. An extension to the traditional auction paradigm that is relevant to the TSE where more than price is considered, is multidimensional auctions, also referred to as multi-attribute auctions. Any important distinction to make with regards to auctions is that there exist forward or reverse auctions. In the forward auction the seller offers a product to numerous buyers, where the seller “controls” the market because a product is being offered that is in demand by a number of buyers. The price offered by the buyer continues to increase until a theoretical rational market price is met in the market. Supply and demand sets the price. In a reverse auction, the buyer “controls” the market because the item being offered is available from a number of sellers. The price offered by the sellers continues to decrease until a theoretical rational market price is achieved. The basic premise of a reverse auction is that a sufficient supply exists and seller’s profit margins are sufficient to offer reduced prices. The reduced price will be offered because the suppliers can instantaneously observe the prices being offered by other sellers [135, 136].

Multi-attribute (reverse) auctions combine the advantages of auctions, such as high efficiency and speed of convergence, and permit negotiation on multiple attributes with multiple suppliers in a procurement situation. A multi-attribute auction is defined as an item characterized by several negotiable dimensions and first arose in the tenders and procurement area [137]. The advances in information technology also allow the use of varied and more complex auction mechanism, where Fieldman [cf. 111] stated that
"We've suddenly made the interaction cost so cheap, there's no pragmatic reason not to have competitive bidding on everything". If the multidimensional auction has the variable quantity, it is referred to as multiple issue auction.

Laffont and Tirole [138] describes many of the critical issues in procurement negotiations from an economics point of view and also mention the need for a generalization of auction theory to so called "multi-dimensional bidding". Generalizations of standard auction theory to the multi-attribute case has been discussed by Thiel [139], Che [140], Branco [141] and more recently David et al [142,143] and De Smet [144]. One of the most notable contributions is that provided by Che which is further outlined below.

Che [140] discusses the design of optimal multidimensional auctions in government contracting. He investigates sealed bid auctions in two dimensions i.e. price and quality procurement problem. The suppliers' type is modeled by a single cost parameter which is independently and identically distributed across suppliers. The buyer is assumed to know the probability distribution of the symmetric bidder's cost parameters. Che analyzes a first and second score sealed bid multi-attribute auctions as well as second-preferred auction institutions. In each of these auction protocols the buyer/auctioneer announces its item request, which consists of the item's desired characteristics, the auction protocol and a scoring rule for describing its preferences concerning the items properties. The scoring function associates a score with each proposed offer and is used by the auctioneer as a tool for choosing from a set of offers, while it is used by the bidders to calculate the optimal bid. Because the scoring function influences the proposed bid, the buyer agent tries to derive a scoring function that maximizes its own expected utility in a given auction protocol. A seller agent capable of bidding sends its sealed bid which specifies the full configuration it offers. In First-Score Sealed-Bid, each firm submits a sealed bid and upon winning produces the offered quality at the offered price. In Second-Score Sealed-Bid the winner in this auction protocol is required to match the highest rejected score in the contract. The second-score differs from the second-preferred auction in that the latter requires the winner to match the exact quality-price combination of the highest rejected bid, while the former does not have this constraint. Che shows that in the case the scoring rule reflects the buyers true preferences, all three institutions yield the same auctioneers utility. However, Che finds

57
that it is optimal for the auctioneer in the first and second score institutions to manipulate the scoring rule in order to discriminate quality in relation to price.

Within the TSE the multi-attribute auction protocol chosen was First-Score Sealed-Bid. This auction protocol was chosen because of its speed of convergence where service providers submit one offer, instead of bidding over a number of rounds, which reduces the signalling load within the TSE. In addition the auction protocol chosen eliminates the problems that frequently arise in second-price or Vickrey auction protocols where it may suffer from bidder collusion, a lying auctioneer, and undesirable revelation of private information [145], while ascending auction protocols typically suffer from problems relating to collusion and can also deter entry into the bidding since a weaker potential bidder knows that a stronger bidder can always re-bid to top any bid he makes. By contrast, a sealed bid auction provides no opportunity for either signaling or punishment to support collusion. Furthermore, entry is promoted because a weaker bidder knows he has a better chance of winning [146].

Perhaps since multidimensional/multiple issue auctions hold great promise for the improvement of B2B transactions, their development has largely been practice driven. Such practice driven experiments demonstrates that multi-attribute auctions can produce higher gains for participants because of the bidding flexibility it offers [147]. Specifically, Bichler [148] found that the utility scores achieved in multi-attribute auctions were significantly higher than those corresponding to single attribute auctions. A clear conclusion of the study is that in situations of many negotiable attributes, it's necessary to provide bidders with advanced decision support tools, since the determination of the attributes which achieve the highest utility is a rather difficult task.

Similar work was also preformed by Chen Ritzo et al [149] who experimentally compared the multi-attribute auction with the price only action. In order to give the price auction the best chance of success, they calculated the optimal reserve levels of non-price attributes to be announced. This calculation is based on full information about the bidders profit functions in order to provide the most difficult test for the multi-attribute auction to compete with. In spite of this handicap, they found that the multi-attribute auction is still more effective in terms of buyer utility and bidder's profits compared to auctions where price is the only strategic dimension.
3.6 Preference Elicitation

One of the design characteristics of the TSE is that it allows the mobile user to specify their preferences over a host of issues such as performance, rating and price in; and service providers preferences over Web service properties such as performance, rating, availability and price for QoS based service composition. An important issue in this context is how to model the mobile buyer/service provider's preferences in the form of a scoring function for use in the multi-attribute auction, where in practical implementations, the elicitation of buyer preferences, and consequently the construction of the appropriate scoring function are of pivotal importance. A common approach is based on the use of established decision analysis techniques, such as Multi-Attribute Utility Theory (MAUT) [150], Simple Multi-Attribute Utility Theory (SMART) [151], Analytic Hierarchy Process (AHP) [152] or Conjoint Analysis [153].

Although advanced versions of MAUT and AHP can model interactions among attributes, the basic techniques use a linear, weighed value function, which assumes preferential independence of all attributes. An attribute \( x \) is said to be preferentially independent of \( y \) if preferences for specific outcomes of \( x \) do not depend in the value of attribute \( y \)[154]. To formally describe preference elicitation, consider \( I \) bids or offers with \( J \) attributes. Each attribute \( j \in J \) has an attribute space \( K_j \). A multi-attribute offer, received by the buyer, can then be described as an \( n \)-dimensional vector \( v_r = (v_{j_1}, \ldots, v_{j_n}) \) where \( v_j \) is the level of attribute \( j \). In the case of an additive scoring function \( S(v_r) \) for the bid \( v_r \) through a scoring function \( S_j(v_j) \) for a bid \( v_j \) is given by the sum of all individual scorings of the attributes. It is convenient to scale \( S_j \) and each of the single-attribute utility functions \( S_j(.) \) from zero to one. That is, for a bid \( v_r \) and a scoring function that has weights \( w_{j_1} \ldots w_{j_n} \), the overall utility for a bid is given by:

\[
s_r = S(v_r) = \sum_{j \in J} w_j S_j(v_j) \quad \text{and} \quad \sum_{j \in J} w_j = 1.
\]

The problem a buyer faces is to determine appropriate \( S(.) \) functions and \( w_j \) weights. An optimal auction is allocating the deal to the suppliers in a way that maximizes the utility for the buyer i.e. to the supplier providing the bid with the highest overall utility.
score for the buyer. The function \( \max_{i} s_i \) with \( 1 \leq i \leq I \) gives us the utility score of the winning bid and can be determined through open-cry or sealed bid auction schemes.

MAUT is a widely used method for the normative analysis of choice problems. The assessment of appropriate weights \( w_j \) is key to MAUT and is an important aspect of a "good" preference model. However in MAUT, some kind of subjective judgment forms the basis for weights, and yet the interpretation of weights is not always clear. Decision analysis tools like the AHP and Conjoint analysis provide for more sophisticated approaches.

SMART provides a simple way to implement the principles of multi-attribute utility theory without the need for complex software or lottery trade analysis. However the alternatives must be known with certainty in advance and it is hard to derive a more generic utility function from the process. This is the reason why the technique is not suitable for multi-attribute auctions.

Since many decision makers feel unable to provide exact weights, some of the more recent approaches only ask for uncertain estimates. For examples, methods from fuzzy analysis use fuzzy sets for weights and individual scoring functions and fuzzy operators from the aggregation of those fuzzy sets. AHP uses a different approach to weight determination. A principle used in AHP is that comparative judgments are applied to construct a symmetric matrix of pair-wise comparisons of all combinations of attributes. Whereas AHP utilizes ratio scales for even the lowest level, MAUT utilizes an interval scale for the alternatives. While it is difficult to justify weights that are arbitrarily assigned using MAUT, it is relatively easy to justify judgments and the basis for the judgments using AHP. The approach assumes weaker decision makers and do not ask for attribute-level utility assessments [155].

Conjoint Analysis is a versatile marketing research technique, which helps to examine the tradeoffs that people make when deciding on an alternative and can be used to construct additive utility functions. In conjoint analysis, a utility is a numerical expression of the value that consumers place in each level of each attribute. The range represents the maximum impact that the attribute can contribute to a product.
In principle, MAUT, AHP and conjoint analysis allow decision makers to model complex problems in a hierarchical structure that shows the relationships of the goal, objectives and alternative. All of these have been adopted in a wide number of applications. Decision analysis techniques like AHP and conjoint analysis provide more sophisticated approaches compared to MAUT and SMART. A large number of AHP applications have been published. A major reason for this popularity is that AHP is relatively easy to understand, and the overall process provides a rational means of approaching a decision that is very easy to explain to others. It is for these reasons that AHP was chosen as the decision analysis tool in the TSE [111].
3.7 TSE Architecture

As stated in Chapter 1, the TSE is represented an eMarketplace where a market based approach to service provisioning is supported. Within this eMarketplace, eCommerce transactions are conducted over the Internet where the various stakeholders of the system can be differentiated according to the nature of their transactions.

In essence in the TSE, there are two types of transactions. The first type relates to the mobile users being able to exert their bargaining power by purchasing services on a per request basis outside of their subscription contracts from an alternate service provider which are in essence B2C transactions. The second type refers to service providers and their collaboration activities with other service providers which are essentially B2B transactions. To support both forms of transactions within the TSE an exchange was adopted as an architectural component, where an exchange is defined by Collins et al [15] as a collection of domain specific markets in which goods and services are traded, along with some generic services required by all markets. Collins et al defined three fundamental elements to the generalized multi-agent market architecture, including an exchange, the market and the market session. Based on this design decision the TSE's generalized market architecture is shown in Figure 3-1.

![Figure 3-1 TSE Market Architecture](image)

The B2B market within the TSE provides a market where various service providers, MVNOs, and third party service providers and the like, can come together and offer their services and resources to form in dynamic collaborations, with the ultimate aim of creating composite services. These composite services are enabled through a horizontal
approach to service provisioning by the adoption of an SOA framework and Web services that enable them. Adopting the TSE B2B market as a complementary approach to service provisioning with the IMS service environment would enable service and network operators to become fully responsive in the face of new challenges such as technological progress, increased competition and deregulation. The modularity and platform independence of SOA and Web services will greatly improve application development, deployment as well as the risk and cost of undertaking projects. With interoperable software components that encapsulate business functionalities available over the Internet through the B2B market of the TSE, telecom operators are endowed with the flexibility of choosing the best software components by using a well defined preference elicitation and QoS model and pay for the services on an as-needed basis. In the B2C market of the exchange the mobile user can exert their bargaining power where they can purchase services on a per request basis, outside of their subscription contracts.

Within both markets of the TSE the preference elicitation model used was the Analytical Hierarchical Process (AHP) [152] where service providers and mobile users in the B2C market are presented with a series of abstract non technical questions. AHP was chosen as it assumes weaker decision makers and this assumption is important as it has been noted in [8] that mobile users should not be bothered, mainly with the questions of QoS, because the majority of them do not even have the knowledge of the related concepts that is QoS. The initiators preferences then form the basis for the selection process in both markets for the best service provider(s) that suit their needs.

Externally, B2B and B2C integration or collaboration is made more cost effective because the network and service providers do not have to set up a separate integration project with each business partner. As a result, business alliances can be created and decoupled on the fly depending on the performance of the contributing service provider, where the agents representing the stakeholders in the system can monitor this performance through the use of stateful Web services [89] and the Web Service Notification (WSN) [91] framework.

Internally, in both markets the stakeholders are represented as agents, whose inherent characteristics make it possible for them to act of behalf on their owners to meet their design objectives, where they are able to opportunistically adopt goals and be both proactive and reactive in their decision making process. The agents within the TSE
include: the Buyer User Agent (BUA), Trusted Intermediary Agent (TIA), Service Provider Agents (SPAs), Network Operator Agent (NOA) and Better Business Bureau Agent (BBBA). The internal architecture of the TSE showing these agents is shown in Figure 3-2 and are also further outlined below. The common services provided in the TSE are provided by the Agent platform and include the Directory Facilitator (DF), Agent Management System (AMS), Agent Communication Channel (ACC) and the Operations and Management (O&M) database. These common services are further described in Section 5.1.

3.7.1 Buyer User Agent (BUA)

The Buyer User Agent (BUA) is a piece of software, installed on the Buyers' device. This device can be a mobile phone, Personal Digital Assistant (PDA), Laptop or Personal Computer (PC). The BUA is capable of generating and responding to Session Initiation Protocol (SIP) requests as discussed in Section 2.2.3 (see appendix A for more information on SIP) and incorporates a GUI to allow the user to select the media it wishes to incorporate into the call, as well as elicitation of the buyers’ preferences over a range of price and QoS attributes. When obtaining the buyers’ preferences the buyer should not be concerned with complicated technical terms relating to QoS, instead in the TSE the buyer is presented with a series of non technical abstract question and as a
result the Analytic Hierarchy Process (AHP) has been selected as the means for preference elicitation in the TSE.

3.7.2 Market Interface Agent (MIA)

The Market Interface Agent (MIA) represents the entry point between the Buyer User Agent (BUA) and the markets within the TSE exchange. The MIA accepts BUA requests in the form of SIP requests with SDP payloads and appropriately converts them to Foundation of Intelligent Physical Agent (FIPA) Agent Communication Language (ACL) requests, required for intra market communication. Once the BUA SIP request is converted into the appropriate form the MIA forwards the service request onto the Trusted Intermediary Agent (TIA) in either the B2B or B2C market using the semantics and ontology used in the eMarketplace. The MIA also accepts presence and context information from the BUA to maintain the user’s status and location information.

3.7.3 Trusted Intermediary Agent (TIA)

The Trusted Intermediary Agent (TIA) is the agent that trades on behalf of the buyer in the B2C market and acts as the broker required for service composition in the B2B market. It is so called as it is ‘trusted’ by the buyer/service provider to act within the rules set to enable automated negotiation procedures in the TSE. Its main functions include:

- *Initiates automated negotiation:* The TIA goal is to locate a service provider that provides the best bid for the initiating entity. The automated negotiation protocol set for both the B2B and B2C markets is the multi-attribute auction. Using this auction protocol, the TIA uses the preferences obtained from the buyer/service provider to set the scoring rule for describing the entities preferences concerning the service properties. This information is represented using the market ontology and is sent in the content of the *Call for Proposal* request sent to the various service provider’s capable of servicing the buyers request.

- *Prevent Value and Time Based Counterspeculation:* The TIA also formalises a relative time, $t$, by which all SPAs must have their bids submitted. This time, $t$, is used to prevent time-based counterspeculation, while ensuring all bids are sealed prevents value based counterspeculation techniques.
3.7.4 Service Provider Agent (SPA)

The Service Provider Agents (SPAs) are agents operating on behalf of service providers in the market. There is no restriction over what form the service providers may present themselves. A service provider may be a Mobile Virtual Network Operator (MVNO) with a pre-negotiated contract with a network operator, Telecom mobile service provider, Wireless Access Provider or could equally be an individual capable of offering computing resources, a database or value added applications. This approach provides an open dynamic framework, where service providers can choose to participate depending on existing commitments, supply and demand. Similarly they can choose a price for their service based on such criteria. The only requirement for the SPA is that they operate within the rules of the marketplace and use the ontology and interaction protocols set in place for automated negotiation.

One of the SPA’s main functions is that it determines participation in bidding sessions depending on the QoS requirements solicited from the buyer and also calculates the QoS levels and price that it is willing to offer for the request. In addition, the SPA monitors the state of its Web services enabled by the Web Service Notification (WSN) framework.

3.7.5 Network Operator Agent (NOA)

The Network Operator Agent (NOA) is the agent that acts on behalf of the network operator in the market and is responsible for interacting with policy management entities in the operator’s home network. In order to interact with the other agents in the marketplace especially with the SPA, the NOA is capable of generating and responding to SPA requests. In addition the NOA can formulate a wholesale price for a service if responding to a SPA who essentially is a MVNO. Additionally in order to communicate with its home network the NOA is also able to generate and respond to SIP requests.

3.7.6 Better Business Bureau Agent (BBBA)

The Better Business Bureau Agent (BBBA) provides neutral impartial information and repository services to the entities within the TSE. The information services that it provides include the reporting and maintenance of the service providers rating in the market while the repository service enables service providers who have negotiated
terms in the TSE, to store and maintain the SLA as a result of such negotiations. The SLAs can be formed between various service providers in the B2B market in addition to SLAs formed between service providers and mobile users. One of the marketplace rules set by the TSE is that the service providers must report their achieved QoS in relation to the call or service upon completion. The BBBA then checks the service providers negotiated QoS to that actually achieved. The persistent and stateful nature of Web services eliminates the subjective nature of perceived quality and allows agents to effectively participate in this post evaluation stage of the BBT model shown in Figure 2-3 and effectively alleviates the concern raised by He [30] with regards to agent participation in this stage.

3.7.7 WS-Resource Integration Gateway Agent (WSRIGA)

<table>
<thead>
<tr>
<th>Area</th>
<th>FIPA</th>
<th>Web Services &amp; WS-Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service Description</td>
<td>Agent Description Ontologies</td>
<td>Web Service Description Language (WSDL)</td>
</tr>
<tr>
<td>Registration</td>
<td>Directory Facilitator</td>
<td>Universal Description Integration</td>
</tr>
<tr>
<td>Communication Protocol</td>
<td>Agent Communication Language</td>
<td>Simple Object Access Protocol</td>
</tr>
<tr>
<td>Semantic Language</td>
<td>FIPA-SL</td>
<td></td>
</tr>
<tr>
<td>Interaction Schemes</td>
<td>FIPA Agent Interaction Protocols</td>
<td>Business Process Execution Language, Web Service Choreography Description Language</td>
</tr>
</tbody>
</table>

Table 3-1 Agent and Web service communication languages and protocols

Current FIPA agent communication specifications lack interoperability with Web services standards and technologies. Although FIPA has a concrete XML representation of ACL messages and envelopes, it still has not ventured outside the realm of agent-to-agent communication using a few transport protocols (i.e. HTTP, IIOP and WAP). This provides serious limitation as software agents are viewed as the components that will “realise and request Web services” and are considered the “brain” component in enabling WS-Resources. As a result of this limitation IEEE FIPA standard committee called the Agent and Web service interoperability working group has been set up to “fill the communication gap between agents and Web services”, creating an environment
where agents are able to locate, negotiate and interact with Web services seamlessly and visa versa. Similarly, the functional requirements of the TSE require that this communication gap needs also be filled between agents and (stateful) Web services. This communication gap and mismatch of communication protocols are further outlined in Table 3-1.

In response to the above limitation numerous research projects have attempted to overcome the problem of agent and Web services integration using a proxy approach. A proxy based approach allows the two platforms to be evolved in parallel without imposing any restrictions on each other. This approach, advocated by the AgentCities project, accepts the equity between the roles of agents and Web services, which is different to the traditional view that agent platforms are considered one level up from Web services, and agents take solely the roles of Web service providers and consumers.

The “FIPA Agent Service to Web service Gateway” as shown in Figure 3-3 lies on the border between the two environments allow FIPA agents to access Web services by translating ACL messages to Web service innovations. In the reverse order the “Web service to FIPA Agent Gateway” exposes and registers agent services in UDDI registry server so that any Web service client can use them. In the TSE this proxy gateway is referred to as the Web Service – Resource Integration Gateway Agent (WSRIGA), the development issues of which is further outlined in Section 5.2.

Figure 3-3 Integration Architecture [156]
3.8 TSE Negotiation Model

The auctions and their associated strategies in the B2C and B2B of the TSE are similar but inherently different. The B2C market is concerned with the sale of calls and services on a per request basis so, as a result the multi-issue perspective is not relevant here. However the auction does need to consider multiple attributes as specified by the buyer to include details of the price and quality attributes that are important to him/her.

The B2B market in the TSE is concerned with a different set of transactions, more specifically related to B2B collaborations and the formation of Virtual Organisations (VOs). These VOs are composed of a number of cooperating companies that share their resources and skills to support a particular product or service. The formation of a VO involves a selection process based on a number of variables such as organization fit, technological capabilities, relationship development, quality, speed and price [157]. Some kind of broker may assist in identifying the best partners to participate within the VO and facilitates the agents to negotiate with one another in order to set the terms and conditions of their partnership. Then, once the VO and the associated services are composed, the agents need to coordinate their actions so that they deliver their services in an effective manner. Within these VOs a single quantity would be infeasible as the relationships created would be intended to be more long-lived than the transactions that exist in the B2C market. As a result the auction employed in the B2B market can be classified as a multi-issue auction.

As stated in Section 3.5.1 the first-score sealed-bid auction protocol was thought to be the most appropriate, where the buyer/service providers preferences are obtained using AHP as outlined in Section 3.6. However a precondition for the additive utility function used in AHP is mutual preferential independence of attribute values. However, quantity is a special attribute in that it often preferentially depends on price. For example, on financial markets a buyer might be willing to buy 3,000 stocks if the price of a stock is US$29.50, but 6,000 stocks if the price is US$29.

Although quantity is an important issue, it can be shown that the buyer has to use a more complex utility function in only a few cases. For further analysis a classification framework, presented by Bichler [111], has been introduced (as shown in Figure 3-4) which is based on three dimensions:
- **Divisibility of demanded quantity**: Will all the quantity be purchased from a single bidder (i.e. sole sourcing) or from multiple suppliers (i.e. multiple sourcing)?

- **Variability of the demanded quantity**: Does the buyer intend to purchase a fixed quantity of a good or does the buyer’s demand depend on prices offered?

- **Divisibility of bids**: Do bidders accept partial quantities, or do they bid a price for a certain number of items?

![Figure 3-4 Auction types where quality is an issue](image)

In the TSE, quantity is only relevant in the B2B market. The automated negotiation procedures are initiated in the circumstance that a new service needs to be composed. Each atomic service in the composition can be provided by one supplier where they together constitute a new composite service and the service providers comprise a new VO. As a result the broker in the B2B market, purchases a fixed amount of items from potentially multiple bidders i.e. multiple sourcing. In the auction mechanism designed for the TSE price and quantity are preferentially independent resulting in the possibility of divisible bids. This means that the additive scoring function deployed in the B2C market is also relevant here, allowing the strategies devised for the B2C market to be applied in the B2B market with an additional independent attribute of quantity. As a result the TSE using the diagram above is Type B.

Using the above described auction protocols there are three dimensions to the negotiation model devised in the TSE in both B2C and B2B markets. These include: the
mechanism for allowing the user (or service provider) to specify their fuzzy QoS and price (fuzzy or crisp) parameters; the reasoning about these specified parameters by the various service providers; and the selection (winner determination) on the most appropriate service provider based on their requirements. In the presence of multiple service providers with comparable or identical functionality, users will want to discriminate these alternatives based on their desired QoS. However, different applications have different QoS requirements where QoS can encompass a number of non-functional attributes. The applications that are of concern to the TSE are real time traffic services such as voice and video as well as stateful Web services applications.

<table>
<thead>
<tr>
<th>Medium</th>
<th>Application</th>
<th>Degree of symmetry</th>
<th>Data Rate</th>
<th>Key Performance Parameters, target values</th>
<th>Parameters, Info. loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>Conversational class</td>
<td>Two-way</td>
<td>4-25</td>
<td>&lt;150 ms pref. &lt;400 ms limit</td>
<td>&lt;1 ms &lt;3% FER</td>
</tr>
<tr>
<td>Video</td>
<td>Videophone</td>
<td>Two-way</td>
<td>32-384</td>
<td>&lt;150 ms pref. &lt;400 ms limit</td>
<td>&lt;1% FER</td>
</tr>
<tr>
<td>Data</td>
<td>Telemetry</td>
<td>Two-way</td>
<td>&lt;28.8</td>
<td>&lt;250 ms N/A</td>
<td>Zero</td>
</tr>
<tr>
<td>Data</td>
<td>Interactive Games</td>
<td>Two-way</td>
<td>&lt;1 kB</td>
<td>&lt;250 ms N/A</td>
<td>Zero</td>
</tr>
</tbody>
</table>

Table 3-2 3GPP End user expectations of RT services

In order to reason about QoS a model is needed which captures the QoS descriptions of these from a user perspective. This is particularly true with voice where [158] describes conversational traffic as follows “Real time conversation is always preformed between peers (or groups) of live (human) end-users. This is the only scheme where the required characteristics are strictly given by human perception. The maximum transfer delay is given by the human perception of video and audio conversation. Therefore the limit for acceptable delay is very strict, as failure to provide low enough transfer delay will result in unacceptable lack of quality”. In reference to this delay ITU and 3GPP have provided an example of end-user expectations of real-time services shown in Table 3-2 [159].

The most important metric for real time services is delay, where time is a common measure of performance. In real time services performance parameters include delay, delay variation and information loss. It is important to note however that the mobile user in the B2C market does not want to be bothered with complicated QoS terms and concepts [8] and as a result the term performance, when presented to the mobile user, is
used to consider these three key parameters for real time services and Web services. Other parameters that the mobile user can understand that does not require technical competence is reputation (i.e. rating) and price.

In the B2B market service providers initiate the service composition procedures and are assumed to be technically competent over a range of QoS attributes. These QoS constraints are then used in the selection of the service provider and their atomic service to form part of the business process. With the specific area of WS-Resources, Menasce et al [160] defined a number of QoS aspects and metrics, which are adopted in the B2B market. These include:

- **Performance** \( (q_p) \): Given an operation \( op \) of a service \( s \), the execution time \( q_p(s, op) = T_{\text{process}} + T_{\text{Delay}} (s, op) \), where \( T_{\text{process}}(s, op) \) is the time that an instance of a Web service takes to be processed, \( T_{\text{Delay}}(s, op) \) represents non-value-added time such as queuing time needed in order for an instance of a service to be processed.

- **Throughput** \( (q_T) \): is measured in units of work accomplished per unit time. There are many possible metrics depending on the unit of work. At the application layer one may be interested in the number of delayed quote requests per second. At the collective layer one may be interested in the number of queries per second that can be handled by directory services to locate resources across different VOs. Examples of throughput metrics at the resource layer include the effective transfer rate in Kbit/sec.

- **Availability** \( (q_A) \): Availability represents the probability that a service is available and is defined as the fraction of time that a resource/application is available for use. The availability \( q_A(s) \) of a service \( s \) can be computed using the expression \( q_A(s) = T(s) / \theta \), where \( T(s) \) is the total time in which the service \( s \) is available during the last \( \theta \) time.

In addition to the above metrics two additional metrics of price and reputation are important to both markets of the TSE. The importance of price as an attribute is obvious, while the significance of reputation was emphasized in the Kasbah [161] project with the provision of its Better Business Bureau facility.
• **Price** \((p)\): Given an operation \(op\) of a service \(s\), the cost \(p(s,op)\) is the amount of money that a service requester has to pay for executing the service.

• **Reputation** \((q_R)\): The reputation \(q_R(s)\) of a service \(s\) is a measure of its trustworthiness. The value of the reputation is defined as the average rank given to the service by the end users, i.e., \(q_R(s) = \frac{\sum_{i=1}^{n} R_i}{n}\), where \(R_i\) is the \(i^{th}\) end user's rank on a service reputation, \(n\) is the number of times the service has been graded.

In general a quality criterion is defined as a tuple \((P,V,U)\) where \(P\) is the name of the quality criterion; \(V:P \rightarrow V\), where \(V\) is the value of a quality criterion; and \(U:P \rightarrow U\) is the function that gives the unit of measurement for each quality criterion [162]. The quality vector of services presented to mobile users in the B2C markets is defined to be: \(q(B2C) = (q_f(s), q_r(s), p)\) while in the B2B market the quality attributes are: \(q(B2B) = (q_f(s), q_r(s), q_d(s), q_R(s), p)\).
3.9 Conclusion

The purpose of this chapter was to present the TSE generalized and internal architecture bearing in mind the main features that the TSE must exhibit in order to meet the its requirements as outlined in Chapter 1 and 2. Technologies presented in this chapter include Web services and stateful Web services which are required to effectively manage such service instances in a distributed and dynamic environment. Agent technology their applications in addition to semantics and ontologies were detailed in Section 3.3. Automated negotiation was discussed in Section 3.4 outlining the benefit of such in addition to different approaches to enable it, such as auctions and bargaining. Multi attribute auction protocols and the chosen negotiation protocol in the TSE - First Score Sealed Bid, was also discussed in this section. Closely related to automated negotiation is preference elicitation and its techniques such as MAUT, SMART, AHP and Conjoint Analysis were outlined in Section 3.6, where the advantages and disadvantages of each approach are also provided. Finally using the design and architectural designs made, the TSE generalized and internal market architecture was presented along with the TSE negotiation model.
Chapter 4

TSE Operation

As discussed in Chapter 1, in order for service and network providers to adopt the TSE within their networks, the TSE has to be propositioned in a manner where their concerns regarding the technical complexity of implementing the TSE have to be alleviated. The aim of this chapter is to address these technical concerns and how the TSE can operate within their networks. In addition this chapter outlines specifically to the regulating authorities the somewhat minor modifications of the signalling procedures within the IMS to support the TSE and its transactions. This chapter also focuses on the operation and negotiation procedures within the B2B and B2C market of the TSE, specifically outlining how the automated negotiation protocol and service provider’s strategies that can be used to enable the mobile user purchase a service on a per request basis and formation of Virtual Organization’s in the B2B market.
4.1 TSE Mechanism

The following section will describe the means of eliciting the buyers/service providers' preferences over a range of QoS attributes as well as the description of the auction mechanism in the TSE.

4.1.1 Buyers Preferences

The automated negotiation procedures are initiated by the Buyer User Agent (BUA) in the B2C market and the Service Provider Agent (SPA) in the B2B market. The entities in both markets are presented with a preference elicitation GUI. This sample GUI was developed in Java and hence is portable across a range of platforms and devices including mobile devices (J2ME) and is used to acquire the entities' preferences in an appropriate form. The buyer in particular may not have considered its fundamental preferences in sufficient detail and hence must be prompted with a suitable set of questions. The preferences then need to be mapped into a coherent scoring function.

The implementation presented in the TSE uses AHP and an additive scoring function. As already mentioned, the additivity assumption implies that attributes are preferentially independent and there are no interaction effects. AHP [152,111] is based on the mathematical structure of consistent matrices and their associated right-eigenvector's ability to generate true or appropriate weights. In both B2C and B2B markets the mobile user and the service providers are presented with a series of questions in order to determine the ranking of criteria where Figure 4-1 shows the GUI that is presented to the service provider in the B2B market. To aid the buyer in answering these questions combo boxes with the relevant attributes and relative importance (from 1-10) are presented to the user. The meaning of the values from 1-10 range from equally to extremely more important when comparing one criterion to another. For example in the diagram below, the buyer stated that Price is 4 (between moderately and strong) times more important than Performance, is three times (moderate) more important than rating and twice (between equal and moderate) as important compared to availability. Using pairwise comparisons the relative importance of one criterion over another can be expressed i.e. price is 4 times more important than performance fills two entries in the matrix at positions [1,2] and [2,1] with the values 4/1 and 1/4, respectively, as shown in Table 4-1. Squaring the matrix presented below and computing the subsequent
eigenvectors, results in the buyer’s weights over the set of preferences. In the case below these weights are shown in Table 4-2. The assessment of the individual scoring functions and weights is a core issue when using AHP [152].

![Preference Elicitation GUI](image)

**Figure 4-1 Preference Elicitation GUI**

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Attribute Value</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price Weight</td>
<td>0.491</td>
<td>0.491</td>
</tr>
<tr>
<td>Performance Weight</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>Availability Weight</td>
<td>0.134</td>
<td>0.134</td>
</tr>
<tr>
<td>Rating Weight</td>
<td>0.115</td>
<td>0.115</td>
</tr>
</tbody>
</table>

**Table 4-1 Matrix of Pairwise Comparisons**

<table>
<thead>
<tr>
<th>Price</th>
<th>Performance</th>
<th>Availability</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/4</td>
<td>1/1</td>
<td>3/1</td>
<td>2/1</td>
</tr>
<tr>
<td>1/4</td>
<td>1/1</td>
<td>3/1</td>
<td>2/1</td>
</tr>
<tr>
<td>1/4</td>
<td>1/1</td>
<td>3/1</td>
<td>2/1</td>
</tr>
<tr>
<td>1/4</td>
<td>1/1</td>
<td>3/1</td>
<td>2/1</td>
</tr>
<tr>
<td>1/4</td>
<td>1/1</td>
<td>3/1</td>
<td>2/1</td>
</tr>
<tr>
<td>1/4</td>
<td>1/1</td>
<td>3/1</td>
<td>2/1</td>
</tr>
<tr>
<td>1/4</td>
<td>1/1</td>
<td>3/1</td>
<td>2/1</td>
</tr>
<tr>
<td>1/4</td>
<td>1/1</td>
<td>3/1</td>
<td>2/1</td>
</tr>
</tbody>
</table>

**Table 4-2 Eigenvector**

The BUAs in the B2C market or the SPAs in the B2B market send their service allocation request to the Trusted Intermediary Agent (TIA) which can be described as a market auctioneer in the B2C market. The broker is required to select the services to execute in a business process in the B2B market of the TSE and the service providers in the B2C market, so that the user defined utility is maximized and so it is a fundamental element in the transactions in the TSE.

In order for the TIA to rate the quality attributes such as performance, availability, rating and throughput so that the auctioneer can correctly evaluate the responses, a
scaling mechanism is used. This scaling mechanism is based on the Likert scale, which is the most widely used scale in survey research. It is non-comparative, 5 point scale, where data is measured at the interval levels and where numbers indicate the magnitude of difference between items. In the TSE the scale divisions will go from 1-5, leaving an odd number of divisions which is fundamental for allowing the buyer to take a neutral centre value [163].

Using the Likert scale the service providers reputation, service performance and availability attributes will have five discrete abstract quality levels so that:

\[ q_p, q_A, q_R, q_i \in q_i = \{1,\ldots,5\} \]  

(4.1)

Where:

- \( q_p \) value for quality attribute performance
- \( q_A \) value for quality attribute availability
- \( q_R \) value for quality attribute rating
- \( q_T \) value for quality throughput rating
- \( q_i \) all the quality attributes i.e. \( q_p, q_A, q_R, q_T \)

The value 1 represents the worst performance, availability and service provider rating while the value 5 represents the best. As a result the auctioneer perceives the discrete quality levels of each of these non-price attributes as measures of increasing quality.

In the B2C market the TIA is responsible for obtaining a single, invisible service instance (i.e. quantity or \( q_a = 1 \)) from one of \( N \) sellers, where \( N = 1,2,\ldots,N_s \). In the B2B market, the TIA is brokered to obtain more than one invisible service instance, where \( q_a = x \), which is a whole positive number i.e. \( x \in \mathbb{Z}^+ \). The value of \( N \) sellers will vary in each auction round depending on how many service providers have registered a service description for the requested service or constituent service. Once the service request is received by the TIA, the TIA formats a Call for Proposal to the \( N \) sellers capable of dealing with the buyer/service providers’ request.
The bids $b$ received back from the service provider comprises of a technical specification and a price i.e. $(p, q_i)$, where the buyer derives utility from the contract bid $b(p, q_i) \in \mathcal{R}$:

$$U_{BU}(p, q_i) = -p + v(q_i)$$  \hspace{1cm} (4.2)

Where:

- $BU$ Buyer
- $U$ utility
- $p$ price
- $v$ value

In Esmaeilsabzali et al [82] it is described how the buyers and service providers are only concerned with service quality to a certain extent, where it is believed that once a reasonable level of quality is provided, any extra quality would not make the service requester significantly happier. As a result the authors captured this relationship of quality using a natural logarithm function, where the utility of the bid is represented as:

$$U_{BU}(p, q_i) = -p + \sum_{i=1}^{m} w_i \ln(q_i + 1)$$  \hspace{1cm} (4.3)

Where:

- $w_p$ weight for quality attribute performance
- $w_A$ weight for quality attribute availability
- $w_T$ weight for quality attribute throughput
- $w_R$ weight for quality attribute rating
- $w_i$ weights for all the quality attributes i.e. $w_p, w_A, w_T, w_R$
- $m$ number of quality attributes

David et al [142, 143, 164] adopt a similar approach where they used the example of enlarging the speed of a processor from 100 to 200 MHz has a stronger influence than enlarging the speed from 200 to 300 MHz. As a result they represented this quality relationship using $\sqrt{q_i}$, where the effect of $q_i$ is weighed by $w_i$. The TSE follows the
approach (and consequently strategies) provided by David et al [142,143,164] where the utility of the buyer is expressed as:

$$u_{BU}(p,q_i) = -p + \sum_{i=1}^{m} w_i \sqrt{q_i}$$

(4.4)

This function describes the relationship that as the price decreases, the utility to the buyer increases. Given the buyer’s utility function, the TIA or auctioneer will announce a scoring function, which is used for choosing among bids. The scoring function of the buyer may be different from the real utility function as specified by the buyer, in the sense that the announced weights $w_i$ may be different from its real weights $W_i$. In particular, the scoring function is of the form [164]:

$$S(p,q_i) = -p + \sum_{i=1}^{m} w_i \sqrt{q_i}$$

(4.5)

Where:

$S$ scoring function

Where $w_i$ are the weights that the buyer assigns to $q_i$. From the scoring function the announced bid’s value for the buyer is [164]:

$$v(q_i) = \sum_{i=1}^{m} w_i \sqrt{q_i}$$

(4.6)

The announced values of the weights $w_i$ can be equal to or different from the real values of the weights $W_i$. For example $w_i < W_i$, then for some reason the buyer declared a lower utility derived from each unit of $q_i$ than its actual utility $q_i$.

Each service provider in the TSE is able to produce any technical specification in accordance to the buyer request. For each technical specification the supplier will incur a cost when they produce the good for sale in the market. The total cost of producing a good will normally consist of: Fixed Costs, $FC$ which are costs that are fixed irrespective of the level of output e.g. rent on premises, charge to operate within TSE and Variable Costs, $VC$ which are costs that vary with the level of output e.g. each extra unit of a good produced will require additional units of raw materials, labor etc. Total cost, $TC$, therefore is the sum of fixed and variable costs, i.e.
\[ TC = FC + VC \]  
\[ (4.7) \]

These costs can be modeled by a linear cost function, such as \( TC = 20 + 4Q \). From a revenue maximization perspective the most important function with regards to cost is Marginal Cost, \( MC \), which is defined as the derivative of total cost with respect to output. Since \( FC \) are constant, \( MC \) may be shown to equal marginal variable costs, \( MVC \), as follows:

\[
MC = \frac{d(TC)}{dQ} = \frac{d(FC + VC)}{dQ} = \frac{d(VC)}{dQ} = MVC 
\]
\[ (4.8) \]

since the derivative of fixed costs (a constant) is zero. Marginal cost is essentially a measure of how its total cost changes when its output changes. For example when a firm expands production from 80 to 200 bottles a day, its total cost goes up by $12, which gives rise to the marginal cost entry of $12/day)/(120 bottles/day)=$0.10 per bottle. Similarly, the supplier receives revenue when it sells output. The Total Revenue, \( TR \), received is the price of the good, \( P \), multiplied by the number of units sold, \( Q \), that is: \( TR = P \cdot Q \). Marginal Revenue, \( MR \), is the rate of change in the total revenue per unit increase in output, \( Q \):

\[
MR = \frac{d(TR)}{dQ} 
\]
\[ (4.9) \]

Based on the cost function, the quasi-linear supplier's utility function is:

\[
U_i(p, q_i) = \begin{cases} 
MR - MVC = p - c(q_i, \theta), & \text{if } i \text{ provides product} \\
0 & \text{otherwise} 
\end{cases} 
\]
\[ (4.10) \]

Where:

- \( c_p \) cost for producing the quality attribute performance
- \( c_A \) cost for producing the quality attribute availability
- \( c_T \) cost for producing the quality attribute throughput
- \( c_R \) cost for producing the quality attribute rating
- \( c_i \) cost for all quality attributes for service provider \( i \) i.e. \( c_p, c_A, c_T, c_R \)
MR is the price, \( p \), it obtains for the service and MC is the cost of producing the proposed qualities values. If \( MR > MC \) then this indicates that the revenue for accepting the call/service is greater than the cost of supporting it. Similarly when \( MR < MC \) then it would be in the supplier's best interest not to support the call/service, whereas when \( MR = MC \) then the supplier is indifferent between accepting/rejecting the call/service. Marginal Revenue and Marginal Cost is central to the supplier's revenue maximization policy on admission control.

Naturally, service providers would have to afford a greater cost when providing higher quality services. It is therefore necessary to define the cost function of a particular service provider \( i \) as:

\[
c_i(q_i, \theta) \text{ to be increasing in } q_i, \text{ and } \theta
\]

Where:

\( \theta \) \hspace{1cm} \text{private cost parameter}

Within the context of the TSE, the cost refers to admitting another service request while at the same time satisfying the existing service instances on a particular resource. Because WS-Resources are stateful and transient the service provider agent by its nature and characteristics knows the current capacity and utilization of any resource it owns. Based on this information the supplier needs to make a decision whether to accept/reject the service into its network. In making this decision the service provider must evaluate the cost of accepting this new service instance with respect to violating existing Service Level Agreements (SLAs) and accepting the penalty for this violation.

The penalty cost is a variable cost for the service provider, and can be represented using a static penalty function. Many researchers in evolutionary computation have explored variations of distance based static penalty functions [165]. One example from Thangiah [166] uses linear combinations of three constant based penalties for the three constraints of the vehicle routing with time windows problem. Another example described by Homaiffar et al [167] defines \( L \) penalty levels depending on the magnitude of the violation of the constraints, and is a suitable penalty function for use in the TSE. With regards to service performance the service provider through a SLA is contracted to provide a specific level of QoS which can have five discrete abstract quality levels,
If the buyer purchased a QoS at level 5 i.e. $L_{pur}$, and received QoS equivalent to level 3 i.e. $L_{rec}$, then the penalty that the service provider would be a price reduction by some factor to include the violation from level 5 to 4, and 4 to 3, or the sum of penalties from $L_{pur}$ to $L_{rec}$. This penalty function, $P(x)$, can be represented as:

$$P(x) = \begin{cases} 0 & x \in M \\ \sum_{i=L_{rec}}^{L_{pur}} R_{ij} & x \not\in M \end{cases}$$

(4.12)

Where:

- $R_{ij}$: penalty coefficient of the $j^{th}$ constraint and $i^{th}$ violation level
- $m$: is the number of constraints
- $L_{pur}$: purchased level of quality
- $L_{rec}$: received level of quality

Using the rule of MVC a cost is only incurred when the network utilization greater than its maximum capacity, and hence the only relevant cost that is associated with the acceptance a new call. Based on this cost parameter the sellers cost function, $c_i$, [164] is:

$$c_i(q_i, \theta) = \begin{cases} \theta \left( \sum_{i=0}^{n} a_i q_i \right) & \text{if capacity is exceeded} \\ 0 & \text{otherwise} \end{cases}$$

(4.13)

Where:

- $a_i$: coefficient and $a_i > 0$.

### 4.1.2 SPA Strategies

In the B2C and B2B markets of the TSE, the agents operating within the eMarketplaces have a number of pricing options available including: a fixed or flat price model; competitive economic models; usage time; usage period and duration; demand and supply; foresight and loyalty based; historical data model; advance agreement; calendar based and bulk purchase. Simple pricing models such as charging a fixed price is not very effective when users place QoS demands that vary with applications and time [168]. As a result the TSE is using a competitive economic model based on game theory.
and mechanism design and follows the strategies devised by David et al [164], which allow SPAs to maximize their revenue and minimize their costs.

In the single attribute auction protocol, in which price is the only bidding strategy, the bidder should decide how much to bid considering its beliefs about the other competitors. However in the case of the multi-attribute auction the SPA or bidder has to decide about the values of all the quality attributes in addition to price. One may think that the decision about these values should also be influenced by the bidder’s beliefs about the other competitors. However, a lemma provided by [140, pp. 671 ] described how the values of the quality attributes are determined by the bidder are independent of its beliefs about the other competitors. Thus, the only components, that influence this decision are the bidder’s cost parameter and the announced scoring rule.

Representing the game as a static Bayesian game, formally the service provider’s best strategy to solve for its optimal quality attributes (i.e. its private information on the value of the qualities minus its cost of producing it), $q^*_i(\theta)$ for all $[\theta, \bar{\theta}]$ (where $\theta$ is uniformly distributed between $[\theta, \bar{\theta}]$) is:

$$q^*_i(\theta) = \arg \max_{q^1} \{v(q_1, ..., q_n) - c_i(q_1, ..., q_n, \theta)\}$$

This strategy shows how each bidder will decide about the quality dimensions of a bid, given the announced scoring rule and given the bidder’s beliefs about its cost parameters. Using this strategy, Figure 4-2 describes the effect of weights on the quality offered where it is visible that as the weights increase so does the quality value, whereas Figure 4-3 shows the effect of cost on quality where as the cost of providing the service increases the quality offered decreases.
For determining the optimal price to be offered by a bidder in the first-score sealed bid auction the general equation provided by Che [140] and the extension provided by David [164] who considered multiple dimensions is used. Representing the problem as a Bayesian game, the service provider’s strategy for calculating the price to be offered is:

\[
p^* (\theta) = \left( \sum_{i=1}^{m} \frac{w_i^2}{4 \alpha_i} \right) \left( \frac{1}{\theta} + \frac{1}{(\theta - \theta)^{N-1}} \right) \int_{\theta}^{\theta - \theta} \frac{\tilde{\theta} (\tilde{\theta} - \tilde{\theta})^{N-1}}{t^2} dt \tag{4.15} \]

The \( p^* (\theta) \) strategy will enable the seller agent to decide about its bid according to its private cost parameter, the scoring rule and its beliefs about the other sellers where. For example if there are more bidders its price will decrease since the competition increases.
among the sellers/bidders. Therefore the price that each seller demands decreases following the principle of supply and demand. As the supply increases and the demand is constant the prices decreases. As the announced weights \( w_i \), where \( i \in [1..m] \), increase the quality of the proposed item concerning, \( q_i \), increases, then the price \( p^* \) of the bid will increase too, the relationship of which is shown in Figure 4-6. As the private cost parameter \( \theta \) increases, that is, the seller’s efficiency decreases (shown in Figure 4-5) it will suggest lower quality items (shown in Figure 4-2). This can be inferred from the formulas of \( q_i^* \) and \( p^* \): as \( \theta \) increases the denominators of \( q_i^* \) increase so the values of \( q_i^* \) decrease. However since a given seller has to compete against other sellers, it will also suggest lower prices, \( p^* \) (shown in Figure 4-4) when the quality of an item is lower.

![Figure 4-4 Affect of number of service providers (n) on price (when \( w, = 0.4, \theta = 0.6, \bar{\theta} = 1.2 \))]
Figure 4-5 Effect of rising cost on price (when $w = 0.5, \bar{\theta} = 4, n = 2$)

Figure 4-6 Effect of rising weights on price (when $\theta = 0.4, \bar{\theta} = 1.2, n = 2$)
4.2 Initialisation of the TSE

This section outlines the preliminary operations that need to occur before the TSE can function properly. These operations consider primarily the relationship that exists between the service provider, agents, services and the service descriptions, and are important as the TSE cannot operate unless service providers offer services for sale and the agents are created and incorporated with sufficient intelligence to operate on behalf of its owner and their respective services. These operations are further outlined below and are also shown in Figure 4-7.

4.2.1 Create and Register Service Description

1. Create service description for owned services (SPA)
   
   Once the service provider's operating within the TSE has created their agent and incorporated within them a theory of agency model which specifies what actions an agent can or should perform in various situations, (see appendix C.1.2 for more information on the theories of agencies) they must then create their service description for all the services that they are willing to offer for sale in the eMarketplace. Descriptions must contain a name for the service, the type of service and the name of its owner. For example if a service provider can provide a translation service translating Strings from English to German, then the service provider could name it as "translation service" set the services type to "English
"German" and the owner to "Vodafone Ireland". For each service the agent needs to create a service description and place it in an array.

2. Publish a service description as Web service endpoint (SPA)

In order for the agent to interact with the Web services it owns it needs to become a Web service endpoint. To do this the agent creates a service description and sets the type of this description to "ws-resource" and sends the request to the Directory Facilitator (DF), which is the yellow directory pages service in the TSE.

3. Register service description (SPA-DF)

Once all service descriptions are created the agent then registers its agent description with its agent identifier, service description array, along with a list of protocols, ontologies and languages it supports to the Directory Facilitator (DF) in the TSE. The service provider agent at this stage can create a goal to keep the agent registered, preventing the agent from manually registering itself after a lapsed period of time.

4. Deal with registration (DF-WSRIGA)

Upon registering with the DF, the DF inserts the SPA service descriptions in its yellow pages for advertisement to allow search and discovery operations for other agents in the TSE. Upon registration, if any service description has the type set to "ws-resource", then the DF will use this type specification as a trigger to inform the WSRIGA.

5. Performs translation (WSRIGA)

Upon receipt of this request the WSRIGA:

a. Converts the agent service description into UDDI tModels and FIPA ACL message into a SOAP message. The ACL-RDF/tModel-SOAP Codec performs this operation which is a complex bidirectional module that is responsible for parsing ACL messages received from the TSE DF ands extract the RDF service descriptions held within their context.

b. It then translates the RDF service description into a UDDI tModel and returns the result back to the Gateway element.

6. Register in UDDI (WSRIGA – UDDI)

[1] A tModel in the Web service architecture is a data structure representing a service type in the UDDI green pages directory where each business registered with UDDI categorises all its Web services according to a designed list of service types.
Once returned the gateway agent registers the tModel and WSDL in UDDI and a stub is automatically created for the agent service on the Axis server. The agent can now be invoked as a Web service endpoint.

### 4.2.2 Subscribe to services

Now that the agent is exposed as a Web service endpoint it can now subscribe to its stateful Web services using the WS-Notification (WSN) framework as described in Section 3.2. In the description provided below it is assumed that the service in question is again a translation service capable of translating words from French to English and visa versa. This service is exposed as a WS-Resource, and has a number of properties such as: number of clients online (NoOfOnlineClients), the capacity of the service and its utilization (SystemCapacity, SystemUtilization) and the performance of the translation (PerformanceOfTranslation). Of course the WS-Resource may have a number of additional properties relating to throughput etc. but for the purpose of this discussion the above properties are sufficient.

Using the above service description the agent needs to become a NotificationConsumer mainly to two properties of the Messaging and translation service - NoOfOnlineClients and PerformanceOfTranslation. These properties are important to the agent in determining whether it can accept the service request and what level of QoS it can offer to the buyer. Internally the SPA at any time has beliefs over the state of the resource, with regards to these properties and has defined five threshold limits associated with the service performance. These threshold limits enable the simple mapping of performance into levels 1-5 as per the rules of the auction mechanism within the TSE. Changes in the beliefs shown in Figure 4-8 will cause the execution of various plans (and goals) to deal with such changes i.e. if the belief noOfOnlineClientsForMsnTranslate exceeds 1000 then a plan can be initiated to calculate the cost of dropping a client to support another or the plan could ensure that the agent automatically generates a refuse ACL message upon receipt of any Call for Proposals, until such time that the belief noOfOnlineClientsForMsnTranslate falls below the specified threshold once again.

```
<belief name="noOfOnlineClientsForMsnTranslate" class="int"/>
<belief name="performanceOfMsnTranslate" class="Double"/>
```

Figure 4-8 Belief Base of SPA
Based on the above, the following actions occur for the agent to become a notification consumer and also to assist in its decision making process upon receipt of the Call for Proposal from the TIA. These actions are also shown in Figure 4-9.

1. **Subscribe (SPA-WSRIG)**

The SPA needs to become a NotificationConsumer to its MSN translation service. It is assumed that the SPA is already exposed as a Web service endpoint and has an endpoint reference. The SPA now sends an ACL message to the WSRIGA, with the performative set to *subscribe* and the content of the message including the identity of the service to be invoked and the properties that the service provider want to be subscribed to.

2. **Translation Service (WSRIGA-WS-Resource)**

The WSRIGA gateway element upon receipt of this request translates the FIPA ACL Subscribe request into SOAP. The SOAP header includes the WS-Addressing information of the Producer, while the body contains the actual Subscribe message along with the endpoint reference for the consumer. In the case of performance the SPA wants to be notified only if a particular threshold is reached. To enable this the SOAP message also includes a number of preconditions that state whenever a threshold is reached a Notify message must be sent in response. The subscribe request can also contain an application specific subscription policy element that
enables the consumer to specify information such as the frequency of messages etc. The SOAP message is then sent to the producer. A temporary endpoint is set up on the Axis server, which is removed once a response is received and has been passed to the Gateway parsers for return as a FIPA Notify or Agree to the original sender agent.

3. Subscribe Response (WS-Resource-WSRIGA)

The response from the translation service upon receipt of the Subscribe message contains the actual Subscription WS-Resource. The header will contain information for the NotificationConsumer, while the body will contain the endpoint reference for the actual subscription. This response is sent to the endpoint in the WSRIGA. The WSRIGA now needs to convert the translation service response back into an FIPA ACL message with the preformative set to Agree.


From this point on whenever a change in properties occurs that affect the threshold limit stated in the Subscribe SOAP request the producer will sent a Notify SOAP message back to the Web service endpoint.

5. Translation Service (WSRIGA-SPA)

The WSRIGA now needs to convert the translation service response back into an FIPA ACL message with the preformative set to Inform, with the payload indicating the change of properties. The agent upon receipt of this information can update its belief base regarding the state of the service and is now equipped with sufficient information to deal with the buyers request.
4.3 B2C Market Operation

Operations within the B2C market are commenced upon the formation of the buyers' request. Once constructed the main operations and message sequences within the TSE are initiated. The complete B2C eMarketplace operations can be further described under the following headings with a complete diagram of all operations shown in Figure 4-18. These operations are also discussed separately below under the following headings:

1. Initial Buyer’s Request
2. TIA Assignment
3. Call for Proposals
4. Calculate and Return Bids
5. Winner Determination
6. Connection Termination

4.3.1 Initial Buyer’s Request

There are a number of actions that the buyer must successfully complete before they can participate in the TSE, the operations of which are shown in Figure 4-10 and are outlined below.

1. Data Connection Setup

In order to participate in the TSE, the BUA mobile device needs to obtain an IP address from its current network operator, obtained in a procedure known as Data Connection Setup. There are two stages to Data Connection Setup, Attach and Packet Data Protocol (PDP) Context Activation [18].

Once these procedures have been completed the BUA will have a valid IP address and will be able to initiate communications with the TSE, where the mobile user
will have to pay for the data traffic sent between its own device and the TSE. Looking at the SIP messages sizes as specified in [83] the total traffic expected to be sent between the mobile user and the TSE is approximately 2.5 kB, where Vodafone Ireland charges 2 cent per kB resulting in a charge of about 3 cent. However more novel methods of payment for 3G enabled data services are emerging where the industry is looking at bundles where the mobile user’s monthly fee includes a subscription to 3G data services. In this case the traffic sent will be included in their subscription fee and they will not be charged for the signalling sent to the TSE.

2. SIP REGISTER (BUA-MIA)

Now that the buyer is able to communicate with the outside world it can avail of the services and facilities that the TSE offers. To do this the buyer interacts with its SIP J2ME BUA application on their mobile device. Once the application is initiated it will automatically send a SIP REGISTER [83] request to the MIA of the TSE. The REGISTER request essentially associates the buyer’s SIP URI with the machine or IP address into which he is currently logged enabling the TSE to be informed the current location of the buyer. Upon receipt of this information, the MIA sends back a 200 OK response, indicating that the action was successfully received, understood and accepted.

3. SIP OPTIONS (BUA-MIA)

Upon receipt on this response the buyer application can send an OPTIONS request to the MIA. The purpose of this request is two fold from the buyer’s point of view. First of all it allows the buyer application to query the capabilities of the MIA and it supported methods, content types, extensions, codec’s before actually establishing a connection in the form of an INVITE request to the MIA. In addition the returned response from the MIA can return the media capabilities of the TSE encapsulated using the Session Description Protocol (SDP) [84]. Within the context of the TSE, the SDP essentially contains a list of applications and media that the buyer can purchase in the TSE. The media capabilities described in the SDP will not include standard applications such as voice, video, ftp, Web browsing, email etc. but will include the session and media descriptions of any newly composed services that the buyer’s device can support. The buyer upon receipt of this response will insert the session description information into a list presented on the buyer’s mobile device.

4. SIP INVITE (BUA-MIA)
The buyer now decides that they wish to purchase a service from the TSE. To purchase a service the buyer needs to provide details such as the callee (if relevant) and a description of the service that it wishes to pertain. In some situations it may be required that the buyer needs to download additional software to correctly interact with a particular application. These requirements can be specified in the session description of SDP in the u header pointing to a URI of a download page. In addition the buyer needs to specify its preferences over a range of properties related to the service request such as price, performance and rating of the service provider. These preferences are obtained in accordance to AHP and its preference elicitation model in line with that discussed in Section 3.8.

Once the buyers' weights are determined it now needs to be able to represent these preferences for the service it wishes to purchase in the SIP INVITE request sent from the buyer and the MIA. Even though SIP is extensible, it was thought more appropriate to describe the buyer's preferences by extending the attributes in SDP, as the preferences as essentially associated with the session. Attributes in SDP are used as the primary means of extending SDP and can be defined to be used as "session-level" attributes or "media-level" attributes, or both. Media level attributes add information about a media stream, while session level attributes convey addition information that apply to the communication as a whole rather to an individual media stream. As there is no defined parameter in SDP for portraying the price, performance or rating preferences, the TSE proposes to use session level attributes values in the form of:

\[
\begin{align*}
    \text{a=price:<Weight>} \\
    \text{a=performance:<Weight>} \\
    \text{a=rating:<Weight>}
\end{align*}
\]

Figure 4-11 Session level attributes

To ensure these session level attributes are recognized by the global community, it is possible to register them with the Internet Assigned Numbers Authority (IANA) [169]. The buyer now sends its correctly formatted Invite and SDP payload to the MIA.
4.3.2 TIA Assignment

Upon receipt of the Invite request from the BUA, procedures are commenced within the B2C market of the TSE to automatically negotiate the terms of the buyers’ service request. In order to commence these procedures the MIA needs to assign the BUAs request to the TIA. As a result the MIA performs the following actions, which are also shown in Figure 4-12.

![Diagram of TIA Assignment](image)

**Figure 4-12 TIA Assignment**

1. Check for service description (MIA-DF)
   
The MIA checks the yellow pages of the DF to ensure that a service description exists. To do this the MIA sends a request with the service description and the DF returns a list of service providers capable of dealing with such a request in a inform response.

2. The DF returned response is now checked by the MIA resulting in either one of two actions occurring, which are outlined below.
   
a. 100 Trying (MIA-BUA)
   
   If the returned list for the DF indicates the existence of a service provider capable of dealing with the buyers request then the MIA returns a SIP 100 Trying response back to the BUA, which lets the BUA know that the request was received and is continuing to process the request.

b. 503 Service Unavailable (MIA-BUA)
   
   In the case where the DF response indicates that there is no service provider of dealing with the buyers’ service request then the MIA has to terminate the
transaction with the BUA. It does this by sending the BUA a SIP response set to 
503 Service Unavailable.

2. Request (MIA-TIA)

The MIA now uses the FIPA Request Interaction Protocol [170] in the 
communication session with the TIA. Interaction Protocols indicate to agents how 
they ought to go about achieving particular outcomes with other agents. The MIA 
now uses the BUA SIP INVITE request and the marketplace ontology to 
appropriately describe the concepts and terms which is then encoded using RDF(S) 
and encapsulated in the payload of the request message sent from the MIA to the 
TIA. The agent action in the payload is set to sell with the details of the buyer’s 
service request including its preferences as stated in the SIP Invites’ SDP session 
level attributes. The MIA also sets the protocol header to fipa-request, indicating 
to the recipient that the request interaction protocol and its related message 
sequences are to be used in all communications.
4.3.3 Call for Proposals

The TIA essentially is the auctioneer in the marketplace, responsible for initiating the multi-attribute auction and determining which bid (and service provider) is the best based on the scoring function and weights determined by the BUA (i.e. winner determination). To achieve these goals the TIA performs the following actions which are also visible in Figure 4-13.

![Figure 4-13 Call for Proposals](image)

1. Unavailable to deal with MIA Request (TIA-MIA)
   The TIA determines if it can deal with the request. There may be times that the TIA is overloaded with buyer’s requests and to ensure agent efficiency it may decide not to accept it. In this situation, in accordance to the interaction protocol set the TIA would respond with an ACL message with its preformative set to `refuse` back to the MIA. When the MIA receives this message it may search the DF for another TIA that may exist in the marketplace and can reinitiate communication it.

2. Available to deal with MIA Request (TIA-MIA)
   If the TIA is able to deal with the MIAs (and BUA) request then it will automatically generate an `Agree` ACL message to it.

3. Call for Proposal (TIA – SPA)
   The TIA at this stage examines the content of the request for the type of agent action that is required of it. If the action agent in the payload is set to `sell(service, buyer, service providers)` then the TIA realizes it must commence an auction on behalf of the `buyer` to purchase a `service` from a specified list of `service providers`. All the information to support the internal transactions of the agent is specified in the marketplace ontology, encoded using RDF and sent in the `request` payload. To
commence the reverse first-score multi-attribute auction the TIA needs to formulate
the scoring function. As stated earlier in Section 4.3, the scoring function associates
a score with each proposed offer and is used by the auctioneer as a tool for choosing
from a set of offers, while it is used by the bidders to calculate the optimal bid. The
scoring function is in the form $s_i = S(v_i) = \sum_{j \in J} w_j S_j(v_i)$ and $\sum w_j = 1$. For the
purpose of the following discussion it is assumed that the scoring function is
$-p + 0.71q_p + 0.29q_s$, which corresponds to the utility of the buyer as defined in
equation (4.2 and 4.4). Once the scoring function is determined the buyer creates a
new agent action $\text{cfp(service, scoring function)}$ using the ontology and sends a Call
for Proposal message using the Contract Net Interaction protocol [171] to all service
providers in the MIAs request. The protocol in the message set is set to fipa-
contract-net, and a deadline is set by which all responses must be received.
4.3.4 Calculate and Return Bid

At this stage the SPA has received the Call for Proposal (CFP) from the TIA. The SPA can either be a service provider who owns and operates its own network infrastructure or an MVNO. In the latter case the SPA cannot determine itself if it can support the service without contacting its underlying network. In this situation the SPA proxies the received CFP to its corresponding NOA. There are two main aspects to this stage in the B2C market operation. The first stage needs to determine if the network can provide a UMTS bearer service to support the call, while the second stage needs to determine the capacity or utilization of the actual service/resource. These stages can occur in parallel and are further outlined below.

In order to provide End-to-End Quality of Service in UMTS and to address the first stage as outlined above, 3GPP has devised an architecture, which is shown in Figure 4-14. Within this architecture, the GGSN acts as the IP Policy Enforcement Point (PEP). Policy based [172] admission control (and is shown as the gateway in Figure 4-14) ensures that the resources that can be used by a particular set of IP flows are within the “authorized resources” specified in the Go interface. The two main architectural elements for this policy control are the Policy Enforcement Point (PEP) and the Policy Decision Point (PDP). In UMTS, the PDF makes decisions in regard to Service Based Local Policy (SBLP) using policy rules and communicates these decisions to the IP Base Station (BS) Manager in the GGSN, which is the IP Policy Enforcement Point (PEP). The PDF in UMTS makes policy decisions based on information located in the Application Function. The Application Function offers services that require the control of IP bearer resources and maps QoS application level parameters (SDP) into policy setup information, and sends this information to the PDF via the Gq interface. In UMTS, the Application Function and the PDF is the P-CSCF, which is in the same domain as the GGSN. For each authorized use of resource the PDF generates an authorization token upon the request from the Application Function. The authorization token generated conforms to IETF RFC 3313 [173]. The PDF also enables coordination between events in the SIP session level and resource management in the bearer level. The binding mechanism associates the PDP context bearer to the IP flow to support SBLP policy enforcement in the GGSN and can be used in context activation and modification procedures.
The following procedure presumes the NOA Network Infrastructure is a UMTS network with Service Based Local Policy (SBLP) employed in the network and the IP network backbone is DiffServ enabled. The procedure makes reference to Figure 4-14, which shows the UMTS End-to-End QoS Architecture. The UMTS network operations are shown in Figure 4-15 while Figure 4-16 shows the FIPA ACL message operations.

1. **INVITE (NOA – P-CSCF)**
   
   The SIP Invite along with its SDP is reformatted from the information sent in the ontology and sent from the NOA to the P-CSCF located in the IMS of the UMTS network, to decide if the media characteristics of the call are within the SBLP. Because SBLP is used, the Proxy Call State Control Function (P-CSCF) will forward the SDP to the Application Function (AF).

2. **Policy Setup Information (AF-PDF)**
   
   Upon receipt of this request the AF maps QoS-related application level parameters into policy setup information, and sends this information to the Policy Decision Function (PDF) via the Gq interface.

3. **Authorization (PDF-AF)**
   
   If the IP QoS resources are consistent with operator policy rules defined in the PDF, the PDF sends an authorization token in the authorization acknowledgment message to the AF. The authorization token is then forwarded to the NOA.

---

**Figure 4-14 QoS Management Function**

The following procedure presumes the NOA Network Infrastructure is a UMTS network with Service Based Local Policy (SBLP) employed in the network and the IP network backbone is DiffServ enabled. The procedure makes reference to Figure 4-14, which shows the UMTS End-to-End QoS Architecture. The UMTS network operations are shown in Figure 4-15 while Figure 4-16 shows the FIPA ACL message operations.

1. **INVITE (NOA – P-CSCF)**
   
   The SIP Invite along with its SDP is reformatted from the information sent in the ontology and sent from the NOA to the P-CSCF located in the IMS of the UMTS network, to decide if the media characteristics of the call are within the SBLP. Because SBLP is used, the Proxy Call State Control Function (P-CSCF) will forward the SDP to the Application Function (AF).

2. **Policy Setup Information (AF-PDF)**
   
   Upon receipt of this request the AF maps QoS-related application level parameters into policy setup information, and sends this information to the Policy Decision Function (PDF) via the Gq interface.

3. **Authorization (PDF-AF)**
   
   If the IP QoS resources are consistent with operator policy rules defined in the PDF, the PDF sends an authorization token in the authorization acknowledgment message to the AF. The authorization token is then forwarded to the NOA.
4. Service Request (NOA – SGSN)

The NOA now needs to determine if the multiplexing elements in the UMTS network can accept the call [158]. The NOA will as a result have to send a service request to the SGSN with the authorization token.

5. Call Admission Control (SGSN-UTRAN/GERAN)

The SGSN is responsible for mapping the UMTS bearer attributes as defined by the UE in the Invite SDP into RAB attributes and sends a Call Admission Control (CAC) request to the UTRAN/GERAN. The UTRAN/GERAN admission control function determines if it can accept the call and will modify the Radio Access Bearer (RAB) attributes if necessary.

6. Call Admission Control (SGSN- GGSN)

The SGSN will also send a Call Admission Control request to the GGSN. Once again the GGSN determines it can accept the call. If accepted it will send the appropriate response back to the SGSN. The SGSN then informs the NOA of whether the call can be accepted [158].

7. Propose (NOA-SPA)

The NOA now temporary reserves this level of QoS for the BUA, and then returns this QoS along with the authorization token generated by the network to the SPA in a Propose message.

8. Propose Bid (SPA- TIA)

The SPA has now being informed if the network can support the service request by providing a UMTS bearer service. It also has knowledge regarding the state of the desired service by checking its belief base and WSN and can determine. Using this information it need to calculates a price, and the level of QoS attributes that it will offer the buyer. The service provider at this stage has a number of pricing options depending on the strategies employed in the SPA. The pricing options using a fixed or flat price model, competitive economic models, usage time, usage period and duration, demand and supply, foresight and loyalty based, historical data model, advance agreement, calendar based and bulk purchase. Whatever pricing strategy is employed however the design and signalling of the TSE will remain the same, where only the SPA internal logic on how much to charge for a particular service differing. However in choosing a pricing strategy it is important to note that simple pricing models such as charging a fixed price is not very effective when users place QoS demands that vary with applications and time. As a result the TSE has
developed a competitive economic model using game theory and mechanism design to develop optimal strategies that enable it to select the level of QoS attributes and the optimal pricing depending on the scoring function announced by the buyer and its benefits about the number of other service providers capable of bidding for the request. These strategies are previously outlined in Section 4.1.2. Using these strategies the SPA formulates a bid and returns it to the TIA in the payload of a propose response.

Figure 4-15 Network Operations for calculating and Return Bid
Figure 4-16 TSE Calculate and Return Bid Procedures
4.3.5 Winner Determination

At the end of relative time $t$, bids submitted by the various SPAs need to be evaluated. The multi-attribute auction protocol and its scoring function used by the TIA are used to solve the winner determination problem. The scoring function announced by the TIA was $-p + 0.71q_p + 0.29q_R$ as provided in Section 4.3.3, and the buyer utility function specified in equation (4.2). Using this information the process involved in winner determination is outlined below and is shown in Figure 4-16.

![Figure 4-17 Winner Determination](image)

<table>
<thead>
<tr>
<th>SPA Name</th>
<th>SPA Rating</th>
<th>Bid Received</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Price</td>
<td>Performance</td>
</tr>
<tr>
<td>SPA 1</td>
<td>1</td>
<td>0.22</td>
<td>3</td>
</tr>
<tr>
<td>SPA 2</td>
<td>5</td>
<td>0.80</td>
<td>5</td>
</tr>
<tr>
<td>SPA 3</td>
<td>3</td>
<td>0.50</td>
<td>3</td>
</tr>
</tbody>
</table>

**Table 4-3 SPA Bids**

1. SPA Rating (TIA-BBBA)

In order to fully evaluate the bids received from the SPAs, the TIA needs to have a belief set about the SPAs rating in the TSE. These beliefs needs to be monitored using a Maintain goal with a recur delay set to a positive long value, enabling the state or the belief regarding the SPAs rating to be continuously monitored. Whenever the delay i.e. 4 minutes, has being expired then a plan will be initiated to query the BBBA regarding the SPAs rating in the marketplace. The response sent from the BBBA will contain an updated list of the SPA rating in the TSE, which are then used to update the TIA belief set.

2. Evaluate Bids (TIA-SPA)
After relative time $t$ has expired, the TIA now evaluates all bids received from the SPAs. The SPA will return a value for their offered performance selecting an integer value of between 1 to 5, and price as a non-negative real number. The rating for the SPAs that are held in the TIAs' belief set is also in integer form holding a value from 1-5. An example of SPA bids and the computation of their associated score are shown in Table 4-3. Winner determination is achieved by selecting the SPA who has the highest score, and using the example provided above SPA 2 is the winner of the auction round. The TIA now needs to notify the winning SPA that their bid has been accepted and so formats an accept-proposal response to SPA 2 and a reject proposal to SPA 1 and SPA 3. The TIA also inserts the score of the winning bid, which the losing SPAs can use to compare it to the score it offered.

3. Reserve QoS (SPA – NOA)

The winning SPA now needs to reserve the agreed QoS for the BUA, and as a result it forwards the BUA International Mobile Subscribers Identity (IMSI) to the NOA, who forwards this to the networks Home Location Register (HLR), enabling the BUA to attach during Data Connection Setup. The BUA can also at this stage be assigned a QoS profile.

4. Store SLA (SPA-BBBA)

The TIA informs the BBBA which SPA was the winner, along with the agreed QoS between the parties.

5. Inform-Done (SPA-TIA-MIA)

The SPA also forwards the authorization token and the networks Network IDentification/System IDentification (NID/SID) to the TIA in the content of an Inform-done message. The authorization can be used by the BUA when establishing a PDP context with the network. The TIA forwards this inform-done message to the MIA who originated the request on behalf of the buyer.

6. 200 OK (MIA-BUA)

Upon receipt of this message from the TIA, the MIA sends a 200 OK response, to the original SIP invite request, and the BUA responds with an ACK.

7. SIP REFER (MIA-BUA)

The MIA now needs to transfer the communication session to the negotiated network operator. The SIP REFER [174] method is used to accomplish this task. The originator of this request is the MIA in the B2C market, the recipient is the
BUA and the final-recipient is the network identified by the NID/SID parameters passed from the SPA.

<table>
<thead>
<tr>
<th>BUA</th>
<th>MIA</th>
<th>DF</th>
<th>TIA</th>
<th>SPA 1</th>
<th>SPA 2</th>
<th>SPA 3</th>
<th>NOA</th>
</tr>
</thead>
</table>

**Section 4.3.1 Initial Buyer Request**

- 2. REGISTER
- 2. 200 OK
- 3. OPTIONS
- 3. 200 OK
- 4. INVITE

**Section 4.3.2 TIA Assignment**

1. Request
2. Inform
2 (a) 100
trying
2 (b) 503
Service Un.
3. Request

**Section 4.3.3 Call for Proposals**

1. Refuse
2. Agree
3. CFP
- 3. CFP(sell/service, buyer, service providers)

**Section 4.3.4 Calculate and Return Bid**

- CFP
- Check Capacity
- Determine Bid
- Determine Bid
- Determine Bid
- 8. Propose
- 8. Propose

**Section 4.3.5 Winner Determination**

- Evaluate Bids
- 2. Accept Proposal
- 2. Reject Proposal
- 3. Accept Proposal (Reserve GoS)
- 5. Inform-done

- 6. 200 OK
- 7. REFER

Figure 4-18 FIPA ACL Message Sequences for B2C Market

107
4.3.6 Connection Establishment

Upon receipt of the SIP REFER method from the MIA, the BUA needs to connect to their new network using system acquisition procedures. In system acquisition the BUA is equipped with the intelligence to obtain the NID/SID of the network operator, from the REFER method, disconnect from the current network operator and connect to the new service provider. Once connected to its new network the mobile device needs to perform data connection setup, as outlined previously in Section 4.3.1, and execute the purchased service using SIP based IMS procedures. These steps are shown below in Figure 4-19. The mobile user could have purchased either two forms of services: a real time service where there is an originator and recipient in the communication or; an application driven service where the setup procedure consists of the originator only. The SIP IMS procedures for these types of services are outlined below in Section 4.3.6.1.

![Figure 4-19 BUA Connection Establishment Procedures]

4.3.6.1 SIP Based IMS Procedures for Real Time Service

If the service purchased was a real time service such as a mobile call that has an originator and recipient, 3GPP specifications define end-to-end session flow for this type of service to consist of three types of procedures: Mobile Origination (MO), S-CSCF-to-S-CSCF, and Mobile Termination (MT). MO procedures specify the signaling path between the mobile device initiating a session setup attempt and the Serving-CSCF that is assigned to perform the session origination service, while MT procedures specify
the signaling path between the Serving-CSCF assigned to perform the session termination service and the UE. The destination endpoint or UE to this call can be:

- **Another IMS Mobile**: that may or may not be located in the same network as the initiating mobile. The two mobiles can set up a session through their CSCFs
- **A phone connected to the PSTN**: By going through the Media Gateway Control Function (MGCF) an IMS mobile can set up a session with a traditional PSTN phone;
- **A multimedia device connected to the Internet**: The destination address may indicate that the destination party is on the Internet. Then the session will be set up through the IMS and the IP network will be the destination. A SIP user agent client or SIP server are examples of a multimedia device connected to the Internet.

In the case where the mobile user purchased a application driven service that was composed used SOA and Web services then the IMS procedures involved for this type of service consists of Mobile Origination (MO) and S-CSCF-to-Parlay X OSA gateway, where the S-CSCF performing service control determines how the service should be handled. The OSA framework in its Network functions [26] (i.e. call and session Control, packet switched session control, IM session control functions and QoS management functions) determined the Parlay X Service Capability Features (SCFs) to deal with service request received from the S-CSCF. The SIP based message sequences are essentially the same for both types of services with the service control in the MO call setup differing.

IP Multimedia Call control procedures for MO and MT based on SIP and SDP are described further below in accordance with 3GPP TS 24.228 [175]. The following assumes that the destination endpoint is another IMS mobile and that the BUAs endpoints have chosen to use the GPRS procedures to guarantee the QoS, which means both the BUAs establish satisfactory PDP context on their respective accesses. It is also assumed that the core network is DiffServ enabled and service based local policy (SBLP) decisions are taken by the PDF.

In addition, it should be noted that the focus of the signalling procedures presented is on the modifications required in MO setup procedures only, as MT setup remains the same because the TSE transactions does not affect it. For further details on MT call setup, see [175]. The rationale behind this is just to highlight to the regulating authorities and
service providers, the minor modifications required in the signalling procedures necessary to support a mobile user purchasing a service on a per request basis.

4.3.6.2 Mobile Originating Call Setup

The below example applies to visiting subscribers when the home network operator desires to keep its internal configuration hidden from the visited network. The BUA is located in the visited network and determines the P-CSCF via the CSCF discovery procedure. During registration the home network allocates a S-CSCF. The home network also advertises the I-CSCF as the entry point from the visited network, who forwards the requests to the S-CSCF. When registration is complete, the P-CSCF knows the name/address of the next hop in the signalling path towards the S-CSCF, the I-CSCF. The I-CSCF receives information in the request, from which it determines the name/address of the proper S-CSCF. In the below description the forward signalling path refers to the signalling path from the originating IMS UE – BUA, to the P-CSCF, to the I-CSCF, to the S-CSCF, while the reverse signalling path refers to the signalling path from the S-CSCF, to the I-CSCF, to the P-CSCF, to the originating IMS UE - BUA. These procedures are outlined below and are further described in Figure 4-20.

1. Invite Request (BUA – P-CSCF)

Now that a primary PDP context has been created, the BUA can send the original SIP Invite request to the P-CSCF, which also includes the authorisation token previously generated by the network. This token is inserted in the P-Media-Authorization header field. Because an authorisation token is included in the SIP message the MO network does not need to authorize the QoS again. Because the BUA already has the authorization token, this part of the signalling procedure is modified and hence outside normal MO setup signalling.

2. Invite Request (P-CSCF – I-CSCF)

The P-CSCF remembers the next hop CSCF for this mobile which is the I-CSCF and so the INVITE request is forwarded through this I-CSCF to the S-CSCF.

3. Service Control (S-CSCF)

The S-CSCF validates the service profile and performs any origination service control required for the subscriber and determines the location of the called party based on the information contained in the To header of the SIP INVITE. The called party responds with a provisional response, 183 Session Progress, that will include a SDP in the message body.
4. 183 Session Progress (S-CSCF – P-CSCF)

The S-CSCF forwards the received 183 Session Progress provisional response back to the P-CSCF which proxies the request back to the BUA.

5. PRACK (BUA-S-CSCF)

The BUA now responds with a PRACK which takes the forward signalling path to the S-CSCF. When the terminating endpoint responds it sends a 200 OK, which proceeds to take the reverse signalling path to the BUA.


The primary PDP is now insufficient to ensure QoS, as this PDP context was related to SIP signalling only. BUA now sends an Activate PDP context message to the SGSN, and once again includes the authorisation token in its request [176] as defined in 3GPP TS 24.008. As the resources are already reserved for the BUA on confirmation that it’s SPA was the winner of the bidding session, the SGSN makes reference to this reservation by looking at the authorisation token in the PDP context request. This part of the signalling process is a modification on normal MO setup procedures.

7. PDP Context Accept (SGSN – BUA)

The SGSN sends an Activate PDP Context Accept message to BUA containing the negotiated value of the UMTS QoS Information Element as defined in TS 24.008 [176].

8. UPDATE (BUA-S-CSCF)

The BUA now sends the UPDATE request to the terminating endpoint, via the forward signalling path established by the INVITE request.

9. Ringing (S-CSCF-BUA)

The terminating BUA endpoint device now is ringing and consequently and as a result a 180 Ringing response is sent from the S-CSCF taking the reverse signalling path to the BUA. The BUA responds with PRACK.

10. Authorise QoS Resources (GGSN-BUA)

When the S-CSCF receives the 200 (OK) response to the INVITE request it forwards it to the P-CSCF. The PDF in the P-CSCF sends a COPS DEC message to the GGSN to enable the use of the authorised QoS resources, i.e. to open the 'gate', and allow packet flow in both directions in accordance with the policy decision within the GGSN Policy Enforcement Point. The GGSN receives the COPS DEC message and enables the use of the authorised QoS resources, i.e. opens the 'gate'
within the GGSN, and sends a COPS RPT message back to the PDF. The 200 OK is then sent back to the originating BUA.

11. ACK (BUA-S-CSCF)

Upon receipt of this message the originating BUA sends an ACK message using the forward signalling path to the terminating BUA.

12. NOTIFY TSE (BUA – MIA)

In addition it also sends a Notify message back to the MIA, informing it of its successful connection to the UMTS network. A BYE request is then sent from the MIA, closing the existing communication session between the BUA and the B2C market. The call is now established between the BUA and the end recipient at the specified QoS through the UMTS network.
Figure 4-20 MO Call Signaling with QoS Guarantees
4.3.7 Connection Termination

The call is now in progress, between the BUA and its intended end-recipient. Once the call has completed the following events occur. These events are shown also in Figure 4-21.

1. UMTS Network - NOA

Once the BUA/End Recipient sends a BYE SIP request, the call has been completed. Once completed the network gathers the achieved QoS and statistics of that call and sends this information to its NOA. The NOA then forwards this information to the SPA in a FIPA Inform message.

2. SPA - BBBA

Upon receipt of this information the SPA informs the BBBA of the achieved QoS, and the BBBA will update its database based on the returned statistics.

3. SPA - Operations and Management

The SPA will also update the accounting information in the Operations and Management Database, so the BUA will be appropriately billed for the call.

Figure 4-21 Connection Termination
4.4 B2B Market Operation

The B2C market of the TSE focuses on the sale of mobile services on a per request basis. The B2B market on the other hand supports the dynamic formation of Virtual Organizations (VOs) with the ultimate aim of dynamically composing services based on a service provider request and QoS specification. The service composition procedures outlined below correspond to the general framework proposed by Rao & Su [177, 178] for automatic web service composition and include the following steps:

1. Creation of a Process Model specifying control and data flow among the activities that need to be created
2. Providing information on how concrete services must be bound to the process activities to be discovered, where the service composer usually interacts with the service registry
3. Availability of the composite service, where it must be made available to potential clients through the use of a service registry
4. Invocation of a composite service using a co-ordinating entity e.g. a process execution engine that manages the control and data flow according to the specified process model [179].

One of the main questions left unanswered in this framework was the selection of the most appropriate plan, if more than one exists. The TSE addresses this by adopting a multi-attribute auction framework for the selection of atomic services and ultimately the formation of the ensuing VO and composite service. Using this generic framework the following operations are supported in the B2B market:

1. Initial Service Providers request
2. Call for Proposals
3. Calculate and Return Bid
4. Process Generation
5. Subscribe to service
6. Advertise and execute new service
4.4.1 Initial Service Providers Request

The SPA agent is now capable of initiating a request to commence the automated service composition procedures. Following the model of Web service composition outlined in [177,178] the SPA needs to generate a process model specifying control and data flow among the activities that needs to be created. As stated previously the TSE contribution does not include the specification of this process and execution model. However there are currently a number of approaches that specify address this issue such as ebXML [180], Business Process Management Language (BPML)[47] and Business Process Execution Language for Web Services (BPEL4WS) [181].

The most popular of these approaches is BPEL4WS [181] standard, which is an XML based language enabling users to describe business process activities as Web services and define how they can be connected to accomplish specific tasks. In addition to the process model, choreography is also another activity relating to Web service composition where examples of such languages include the Business Process Specification Schema (BPSS) [182] and Web service (WS)-Choreography Description Language (CDL) [183]. For these activities to be supported the owning or initiating service provider of the composition process will have to describe the collaboration based process models, where once defined the workflow between the various participants to execute a process from start to finish will have to be defined. Once the atomic Web services are specified in the workflow, the B2B marketplace procedures can be commenced to define which Web service atomic elements actually participate in the business process, where consequently the following procedures take place which are also shown in Figure 4-22:

1. SPA Preference Elicitation

   In AHP judgments are derived from the service provider and are used to determine the ranking of criteria i.e. ranking of price, performance, rating and availability. The procedure for obtaining these preferences is outlined in Section 4.1.1.

2. Formulate Request

   Once the SPA preferences are obtained it represents this knowledge using the ontology within the market. The process model and weights are then sent to the TIA in the payload of a Request Agent Communication Language (ACL) message using the FIPA Request Interaction Protocol [170]
4.4.2 Call for Proposals

The TIA is essentially the auctioneer in the marketplace, responsible for initiating the multi-attribute auction and determining which bid (and service provider) is the best based on the scoring function and weights determined by the SPA (i.e. winner determination). To commence the reverse first-score multi-attribute auction the TIA performs the following actions, which are also shown in Figure 4-22:

1. Searches Directory Facilitator (TIA-DF) and Formulates Response (TIA-SPA)

Upon receipt of the request from the SPA the TIA needs to determine if it can support the SPAs request by querying the Directory Facilitator (DF). For each atomic service in the process model, the SPA sends a request to the DF for a list of service providers capable of providing an instance of the atomic service. At this stage one of two things can happen:

a. *Agree to request:* The TIA has a list of service providers for each atomic service and determines it can support the SPAs request and consequently sends back an *Agree* ACL message to the SPA in accordance to the FIPA Request Interaction protocol.

b. *Refuse request:* An ACL *Refuse* message is sent back to the SPA, if the situation arises that for one or more atomic services in the process model there did not exist a service description and consequently service provider capable of dealing to the request.

2. Create the scoring function

As stated earlier the scoring function associates a score with each proposed offer and is used by the auctioneer as a tool for choosing from a set of offers, while it is used by the bidders to calculate the optimal bid. That is, for a bid $v_i$ and a scoring function that has weights $w_j...w_j$, the overall utility for a bid is given by $s_i = S(v_i) = \sum_{j \in J} w_j S_j(v_j)$ and $\sum_{j \in J} w_j = 1$ [111]. For the purpose of the following discussion it is assumed that the weights determined by the initiating service provider are those presented in Table 4-2. The TIA uses these weights and sets the scoring function to be $- p + 0.50q_p + 0.25q_R + 0.25q_D$, which corresponds to the utility of the buyer as defined in equation (4.2 and 4.4).

3. Send Call for Proposal (TIA-SPA)
The TIA now has a list of service providers capable of dealing with the request for each atomic service and the scoring function over the SPAs preferences. It now creates a new agent action \textit{cfp(service, scoring function)} using the ontology and sends a \textit{Call for Proposal} ACL message using the Contract Net Interaction protocol [171] to all service providers in the returned list. The protocol in the message set is set to \textit{fipa-contract-net}, and a deadline is set by which all responses must be received.

4.4.3 Calculate and Return Bid

Upon receipt of the CFP message from the TIA the following events occur. These procedures are very similar to that presented in the B2C market and are also shown in Figure 4-22:

1. Forward request to NOA

The SPA can either be a service provider who owns and operates it own network infrastructure or an MVNO. In the latter case the SPA may not be able to determine itself if it can support the service without contacting its underlying network. In this situation the SPA proxies the received CFP to its corresponding NOA. This procedure is shown in Figure 4-22 for SPA3.

2. Propose (SPA-TIA)

The SPA/NOA has beliefs of the state of the WS-Resource and has knowledge regarding its commitments. Using these beliefs it will determine whether it is feasible to participate in the service composition process. If they decide to bid then the agent must now decide how much to bid using the strategies outlined in Section 4.1.2 and returns the full configuration of its offer in a \textit{propose} response back to the TIA.

4.4.4 Process Generation

Upon the return of bids for the service providers, or the end of a relative time, the TIA performs the following, which is also shown in Figure 4-22:

1. Winner Determination

The TIA determines the winning SPA for each atomic service, using the scoring function. Using the example presented in Table 4-4 SPA2 derives the highest score and as a result SPA2 atomic service will participate in the business process. The
score is calculated for each atomic service and the composite service is composed using SPAs with the highest scores. Once the appropriate SPAs have been selected for the composite service SLAs is accordance to that negotiated is formed between the various entities and stored in the BBBAs’ database.

<table>
<thead>
<tr>
<th>SPA Name</th>
<th>SPA Rating</th>
<th>Bid Received</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Price</td>
<td>Performance</td>
</tr>
<tr>
<td>SPA 1</td>
<td>1</td>
<td>0.22</td>
<td>3</td>
</tr>
<tr>
<td>SPA 2</td>
<td>5</td>
<td>0.80</td>
<td>5</td>
</tr>
<tr>
<td>SPA 3</td>
<td>3</td>
<td>0.50</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4-4 SPA Bids

2. Notification of success (TIA-SPA)
   The TIA notifies the winner/losers by sending accept/reject proposal messages back to the SPAs in accordance to the FIPA Contract Net Interaction protocol.

3. Process Binding
   The TIA at this stage also needs to bind the concrete services to the process activities to be discovered. The use of the Semantic Web and ontologies play an important role in this stage. As described in Appendix B.1.4 the Web Service Modeling Ontology (WSMO) is envisioned to be the semantic markup language enabling the full range of automation activities related to Web services. The information used for describing a task in WSMO is composed of five fundamental elements: signature, preconditions, post conditions, information invariants and non-functional aspects.

4. Inform-done (TIA-SPA)
   Once the process has been bound the TIA notifies the initiating SPA of its success by sending it Inform-done ACL message.

4.4.5 Subscribe to new Business Process

The SPA needs to become a Notification Customer to each atomic service in the business process. This is necessary due to the fact that the QoS of a composite service is determined by the QoS of its underlying component services [70] so in order for the service operators to agree to SLA they must be constantly be informed regarding the
state of any service in the business process. A more detailed view of this operation is outlined below, and is shown in Figure 4-22:

1. Subscribe to atomic services

   The SPA now sends a ACL message to the WSRIGA, with the preformative set to subscribe and the content of the message including the identity of the service to be invoked and the properties that the service provider wants to be subscribed to.

2. Translate ACL to SOAP

   The WSIGA upon receipt of this request translates the FIPA ACL Subscribe request into SOAP. The SOAP header includes the WS-Addressing information of the Producer, while the body contains the actual Subscribe message along with the endpoint reference for the consumer. In the case of the QoS attribute performance for instance, the SPA may want to be notified only if a particular threshold is reached. To enable this the SOAP message also include a number of preconditions that state whenever a threshold is reached a Notify message must be sent in response. The subscribe request can also contain an application specific subscription policy element that enables the consumer to specify information such as the frequency of messages etc. The SOAP message is then sent to the producer.

3. Change in Property

   From this point on whenever a change in a property occurs that affects the threshold limit stated in the Subscribe SOAP request the producer will send a Notify SOAP message back to the Web service endpoint. The WSRIGA now translates the service response back into an FIPA ACL message with the preformative set to inform with the payload indicating the change of properties. The agent upon receipt of this information can update its belief base regarding the state of the service, which can trigger the execution of an event or goal so it can appropriately respond to such change by either: terminating a contact with a service provider; reinitiating the service composition procedures; updating the service providers rating in the BBBA; or notifying another entity in the home network. The WSN and the application of stateful resources allows the health of the system to become more accurate in addition to the SPA being able to appropriately respond to real time changes associated with the dynamic VOs that have resulted from the service composition procedures.
Figure 4-22 B2B ACL Message Sequences
4.4.6 Service Execution and Availability

A key challenge presented in the design and implementation of the TSE is the integration of the composed services into the network operators' home network. There are several ways in which this can be done, but the procedures outlined below make the assumption that the network operator has a Parlay X OSA gateway, which is connected to an Enterprise Service Bus (ESB), as shown in Figure 4-23.

The ESB is used to address the problem of application integration in the IMS to integrate legacy applications and services with Web services. Its implementation is envisioned to be required as existing systems simply cannot be thrown away, as they contain within them great value to the enterprise. Strategically the objective is to build a new architecture that will yield the value hoped for, but tactically, the existing systems must be integrated such that over time, they can be componentized or replaced in manageable fashion. As a result providing support for existing middleware solutions and protocols is probably one of the more important factors in adopting SOA middleware, due to the fact that most organizations will continue to use existing middleware in their network implementations. There is no commonly agreed definition of an ESB - its functionality can range from being a service broker to a comprehensive SOA architectural implementation. However an ESB must at least provide transport, event, mediation services to allow integration of business units and bridging of heterogeneous platforms and environments. A non-compulsory element on an ESB also include an Orchestration service, which is also outlined below.

- The *Transport Service* within an ESB ensures the delivery of messages among the business processes interconnected via the enterprise bus. Transport also includes content-based routing to direct messages to different destinations. As part of a mission critical environment, these services are transactional, secured and managed.

- *Event Services* provide event detection, triggering and distribution capabilities. In an ESB-enabled event driven SOA, applications and services are treated as abstract service endpoints, which can respond readily to asynchronous events. The SOA provides an abstraction away from the details of the underlying connectivity and plumbing, and the implementations of services do not need to understand protocols or how messages are routed to other services. They simply receive a message from
the ESB as an event, and process the message. The event services provided within an ESB are implemented by the EDA as described in Section 3.2 and Appendix B.1.4.

- **Mediation Services** ensures the necessary protocol matching to integrate heterogeneous systems. This may occur if two different services do not use the same underlying transport protocol, requiring the mediation service to provide the transformation from one protocol to another. The mediation service provided normally by the Java Messaging Service (JMS) specified in JSR194, also offers the capability to transform the content of any message into various payloads, which is a key function for business integration to ensure that the data that transits through the bus is understandable by any process.

- **Orchestration Service**: provides for more sophisticated services process management. The functionality of the orchestration service is actually in line with 3GPP in TS 22.127 [201] Release 7 Open Service Access (OSA) specification which extends the OSA specification to include the provision for a Service Broker. TS 22.127 outlines that a service broker enables Service Selection, Service Provisioning, Feature Interaction and Service Chaining. The concept of Service brokering is the ability to package, provision and supply a set of applications or services onwards to the application server implementing the business logic that requires the use of such functionality within an OSA environment and shall enable the delivery of multiple services in an operator networks in a managed and controlled fashion. Therefore whenever an event occurs, there is a need to ensure that the set of applications or services that may act upon that event are invoked in a manner that does not conflict with any other application or service defined in the provisioned package of applications or services. This definition of a service broker is in line with the description of the ESB requiring the use of EDA and SOA in its physical implementation. The Parlay group also supports the use of ESB with its Parlay X OSA API in it Parlay OSA specification Part 3 and 16 [21].
Figure 4-23 Enterprise Service Bus in Network Operators network

Once the service has been composed using the automated procedures outlined above, the SPA or NOA needs to advertise the service description within its home network and provide support for its execution. These procedures are further outlined below:

1. Advertise the service description

   The SPA/NOA needs to advertise the service description in the Web service registry in the provider's home network, making it accessible to the network via the Parlay X Web service Gateway. Before the service can access the underlying network functionality via the Parlay X OSA, it needs to do the following:

   a. Application-network authentication: The authentication mechanism may be supported by cryptographic processes to provide confidentiality, and by digital signatures to ensure integrity. Once authentication the application will be able to select and access an instance of a framework function or network Service Capability Features (SCFs).
b. **Authorization.** In order to use SCFs, the application must first be authorised to do so by establishing a service agreement with the home network environment. Establishing a service agreement is a business level transaction, which requires the OSA that owns the application to agree terms for the use of the SCF, which can be reached using off-line or on-line mechanisms. After a service agreement has been established between the application and the home network environment domain, the application will be able to make use of this agreement to access the SCFs. The authentication and authorization procedures are defined in 3GPP TS 22.127 [26].

2. **Service Brokering**

   Now that the service is authenticated and authorised in the network providers’ home network it can make use of the underlying IMS functional planes i.e. session and transport layers, through an open and secure gateway using Parlay X OSA. The mobile subscribers now may want to avail of the service, requiring its execution. Execution of a composite service can be thought as a sequence of message passing according to the process model. The dataflow of the composite service is defined as the actions that the output data of a former executed service transfers to the input of a later executed atomic service.

3. **Service Control (S-CSCF)**

   When the composite service is invoked in the network operator’s network, the Serving CSCF forwards the service request to the ESB. The ESB performs the necessary procedures to invoke the service according to the predefined process model stored in its database. It can also utilise the SCFs in the Parlay X gateway or potentially directly interact with the HSS and Operation Support Systems (OSS) within the UMTS network and service planes. As the services are stateful they are capable of generating events upon completion or when it is waiting for a new input. These events can be managed by the Event Driven Architecture (EDA) within the ESB, provided possibly by the JSLEE. Upon completion of the service execution an event can be generated back to the ESB who notifies the S-CSCF regarding the successful/unsuccessful completion of the composite service.
4.5 Conclusion

The aim of this chapter was to outline the automated negotiation procedures required within the B2B and B2C markets of the TSE, as well as the various types strategies that can be incorporated within the service providers agents to facilitate its decision making process in deciding to support a service request. In particular, strategies computed using game theory were presented in this chapter, as it was outlined in Section 4.1 that fixed price mechanisms are not suitable for applications where its QoS varies with applications and time.

In addition, key challenges on how the TSE technically provides its goal of service provisioning are addressed within this chapter with key reference to the integration and signalling challenges that exist inside the TSE and in the network operator’s home network. In particular, Section 4.3.4 outlines the integration and signalling procedures required in the SPA, NOA and their respective networks to determine if it can support the requester’s service request with regards its QoS requirements. Section 4.3.6 outlines the modifications required in Mobile Originating Call setup procedures in the event that the service/network provider was successful in its submitted bid. Section 4.3.4 and 4.3.6 would be of particular interest to standardization bodies such as 3GPP to allow for these modifications in future release specifications. Section 4.4.6 also detailed further integration issues, where services composed in the TSE must be made available within the network operator’s home network. Key findings in the section outlined how some kind of service broker or Enterprise Service Bus (ESB) is required for network operators to reap the benefits of the dynamic service composition procedures that take in the TSE. In addition in order to support the event driven nature of WS-Resources an event service must also be provided in the network operator’s home network which could be potentially provided by JAIN SLEE.

In summary this chapter outlined solutions to the technical and signalling challenges that would be of concern to service and network operators attempting to support the TSE in its service provisioning for B3G. In addition, through critical assessment of the signalling procedures presented and qualitative analysis of the presented TSE architecture and design elements presented in Chapter 3, it is possible to prove the first claim presented in Chapter 1, which stated that the TSE creates a flexible, adaptable and
responsive service provisioning platform by enabling: Mobile users purchase services on a per request basis and; service providers to sell services where their interoperable business functionalities available over the Internet can form part of business collaborations. The next chapter will describe the performance related aspects to this presented technical solution proving that the designed message sequences and architectural design elements as presented in Chapter 2 provide a viable and scalable infrastructure for service provisioning for B3G.
Chapter 5

TSE Implementation and Analysis

Chapter 4 focused on the operational signalling requirements to enable the TSE goal of becoming a flexible, adaptable and responsive service provisioning platform for B3G by enabling: mobile users purchase services on a per request basis and; service providers to dynamically collaborate with each other with the aim of dynamically composing a new service. This focus of this chapter however is to use the designed message sequences as outlined in Chapter 4, and quantitatively evaluate the TSE architecture and design to prove that the TSE is a feasible and viable architecture where the acquisition of a service instance on behalf of a mobile user and the dynamic coalition procedures between service providers can take place in a reasonable time. This is the second claim of this thesis as presented in Chapter 1.

In order to help prove this claim a mixture of computer simulations and real time analysis of the proposed architecture was conducted during the scope of this study. The analysis provides an insight into the performance of the proposed architecture under different configurations to evaluate the scalability of the devised message sequences and to identify potential bottlenecks in the system architecture. Within the architecture the B2B and B2C markets and their respective agents and functionality were implemented using Jadex and JADE and the Web Service Resource Integration Gateway (WSRIG) was also tested to determine its performance in its translation and integration activities.

In order to conduct SIP performance analysis in the UMTS network, the Dynamic UMTS Signalling Simulator, ‘SigSim’, was used to estimate the end-to-end signalling load in terms of number of messages handled per network element and procedural delays. The ‘SigSim’ software simulator skeleton, which emulates UMTS Release 5/6, was firstly developed by Motorola [185]. The dynamic nature of ‘SigSim’ derives from the stochastic modelling of users mobility within a particular environment (rural, urban, suburban and dense urban) as well as users behaviour in terms of accessing different services. The simulator implements a model of cell layout and UMTS network, where
the UMTS network consists of the Radio Access Network (RAN) and the Core Network (PS CN) network elements for GPRS and IMS.

This chapter presents details on the tests conducted and the components of the test bed and SigSim UMTS simulator.
5.1 Agent Platforms

In order to prove that the system actually works the TSE architecture had to be developed on some kind of agent platform. To date, there are a large amount of agent platforms available to the community where Table 5-1, outlines a subset of these platforms and indicates whether it is open source and FIPA compliant. For a fully exhaustive list on agents tools, standards and platforms, see [186].

<table>
<thead>
<tr>
<th>Name</th>
<th>FIPA Compl.</th>
<th>License</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent Development Kit (ADK)</td>
<td>Yes</td>
<td>Commercial license, free research license available</td>
<td>Agent development kit that emphasizes the mobile and security aspects, used in several commercial projects especially for legacy system integration.</td>
</tr>
<tr>
<td>Cougaar</td>
<td>No</td>
<td>Open Source</td>
<td>Offers special support for logistics problems</td>
</tr>
<tr>
<td>Soar</td>
<td>No</td>
<td>Commercial license</td>
<td>Is a general cognitive architecture for developing systems that exhibit intelligent behavior. Achieving human level reasoning and decision making for autonomous systems requires agents that are capable of reasoning through large volumes of knowledge. A key element is the ability to resolve conflicts, solve problems and operate in ambiguous and uncertain situations.</td>
</tr>
<tr>
<td>JACK</td>
<td>No</td>
<td>Commercial license</td>
<td>Leading edge commercial BDI-agent toolkit. It represents a legal successor of PRS and dMars, but uses a intuitive language that extends the Java programming language with certain agent specific keywords.</td>
</tr>
<tr>
<td>Java Agent DEvelopment Framework (JADE)</td>
<td>Yes</td>
<td>Open Source</td>
<td>JADE is an efficient open source agent platform developed by TILAB. JADE is widely used in research as well as in commercial projects and has a very active user and developer community</td>
</tr>
<tr>
<td>Jadex</td>
<td>Yes</td>
<td>Open Source</td>
<td>Framework for the creation of goal-oriented agents following the belief-desire-intention (BDI) model. The framework is realized as a rational agent layer that sits on top of a middleware agent infrastructure such as JADE.</td>
</tr>
<tr>
<td>Jason</td>
<td>No</td>
<td>Commercial license</td>
<td>Fully fledged introspecter for an extended version of AgentSpeak, a BDI agent-oriented logic programming language, implemented in Java.</td>
</tr>
</tbody>
</table>
Developed as part of ongoing research at University College Dublin, that is concerned with the creation of "a cohesive framework that supports a structured approach to the development and deployment of agent oriented applications".

<table>
<thead>
<tr>
<th>Agent Factory</th>
<th>No</th>
<th>Commercial license</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Developed as part of ongoing research at University College Dublin, that is concerned with the creation of &quot;a cohesive framework that supports a structured approach to the development and deployment of agent oriented applications&quot;.</td>
</tr>
</tbody>
</table>

Table 5-1 Agent Platforms

When selecting a platform on which to develop the TSE environment, it was necessary to ensure that the choice supported mentalistic concepts to enable the smooth transition from the design to implementation phase. The second key feature the TSE must possess is the provision of a sophisticated communication infrastructure that complies with communication standards and therefore enables platform interoperability.

When dealing with ideas of mentalist notions it can be stated that no general consensus exists about the appropriate set of concepts. This is reflected by the number of agent language families such as Agent-0 [187], AgentSpeak [188] and 3-APL [189]. However, the AgentSpeak BDI model for the implementation of the TSE was seen as the best alternative as the models in the other platforms assume a time-sliced execution of agents that is inefficient when the agents are idle. The BDI model in contrast supports event-based reactive behavior as well as pro-active goal-directed behavior. In addition AgentSpeak is a popular BDI model with a solid theoretical and philosophical foundation, and has been proven to build successful applications. Looking at these requirements there was only one platform that addressed the needs of the TSE, which was Jadex [190].

Jadex, the abstract architecture of which is shown in Figure 5-1 is a software framework for the creation of goal-oriented agents following the belief-desire-intention (BDI) model (for more information on theories of agencies see Appendix C1.2). The Jadex reasoning engine addresses traditional limitations of BDI systems by introducing new concepts such as explicit goals and goal deliberation mechanisms [191] where the behavior of a specific agent is determined solely by its concrete beliefs, goals and intentions.
To develop agents with Jadex it was required to create two types of files, XML Agent Definition Files (ADF) and the Java classes for plan implementations, as shown in Figure 5-2.

To start an agent, the ADF is loaded and the agent is initialized with beliefs, goals, and plans as specified above. For example, the SPA ADF, as shown in Appendix C.1.3, holds beliefs and belief sets over a number things such as whether it is registered or not, the number of services which it can offer for sale, its own rating, the number of competitors for each respective service, service cost etc. This belief base stores believed facts and is an access point for the data contained within the agents.

Goals in agents make up the agents motivational stance and are the driving force for its actions. The handling and representation of goals is one of the main features for Jadex agents. Unlike traditional BDI systems, which treat goals merely as a special kind of
event, goals are a central concept in Jadex, where there are four main types: Perform, achieve, query and maintain.

- A *perform* goal states that something should be done but may not necessarily lead to any specific result.
- The *achieve* goal describes an abstract target state to be reached, without specifying how to achieve it. Therefore it can try out different alternatives to reach its goal.
- The *query* goal on the other hand, represents a need for information. If the information is not readily available, plans are selected and executed to gather the needed information. For example the SPA may need to query the Directory Facilitator (DF) to determine how many other service providers can provide the same service as itself.
- The *maintain* goal specifies a state that should be maintained once it is achieved

In Jadex, goals are represented as objects with several attributes. The target state of achieve goals can be explicitly specified by an expression i.e. referring to beliefs, which is evaluated to check if the goal is achieved. Attributes of the goal, such as the name, facilitate plan selection, e.g. by specifying that a plan can handle all goals of a given name, where the concrete actions an agent may carry out to reach its goals are described in plans. For instance a service provider may have a goal requiring the service provider’s resources to stay above certain parameters. In this case that one of these parameters becomes violated in the agent’s belief set, then the goal will initiate a series of plans to maintain the resources target condition.

Within the TSE these beliefs, goals and plans are used in the SPA to simulate their services changing costs, update its belief base regarding the number competitors and respond to incoming Call for Proposals. The figures below show the output from the BDI tracer that comes with Jadex, which is a tool inspired by the Ph.D. work of Dung N. Lam [192] working on agent software comprehension with abductive reasoning. The tracer provides basically an interface and means to log the internal state of a BDI agent, and to analyze and visualize the logged information. Figure 5-3 output shows this output where changes in beliefs (i.e. circle marked with B) initiate a plan (i.e. circle marked with P) which cause the generation of an event that are traced within the SPA. For example, in the centre of the Figure 5-3, the agent has a Belief (B) where upon
change of this belief cause the execution of a plan (P), which generates a message event (represented as arrows). Similarly, a goal (G) can also initiate an event.

Figure 5-3 Trace Exploration Graph SPA1

Jadex has been used to build applications in different domains such as simulation, scheduling, and mobile computing. For example, Jadex was used to develop a multi-agent application for negotiation of treatment schedules in hospitals [193]. One important aspect of Jadex is that it is middleware independent and consequently can be loosely coupled with its underlying infrastructure. Due to its platform middleware independence Jadex can be realized on either the Java Agent DEvelopment Framework (JADE)¹ or a standalone experimental adaptor for Diet². JADE was selected as the middleware platform to support the basic services required for the TSE to operate, as it is open source and FIPA compliant and it is well recognized as one of the most stable and well developed agent frameworks. The proactive natures of agents are supported by Jadex in its BDI model. As a result of the above the combination of Jadex and JADE were selected.

5.1.1 JADE Architecture and TSE Ontology

The standard model of an agent platform, defined by FIPA and implemented in JADE is represented in Figure 5-4. JADE is used to provide the common services such as the Agent Management System (AMS), Agent Communication Channel (ACC) and Directory Facilitator (DF) required by any market and provides the agents in the TSE with the socialable element required to communicate with other agents. The model of agent communication in FIPA is based on the assumption that two agents, who wish to converse, share a common ontology for the domain of discourse, so that when agent A communicates with agent B, a certain amount of information \( I \), is transferred from A to B by means of an ACL message. Inside the ACL message, \( I \) is represented as a content expression consistent with a proper content language (i.e. RDF, OWL, SL) and encoded in a proper format (e.g. string). Both A and B have their own way of internally representing \( I \) enabled through the use of an ontology [194]. JADE's support for ontologies or agent knowledge model is based on Java classes, which has the obvious advantage over Strings by reducing the amount of parsing activities. Ontology elements and its relations are described as real Java objects, providing powerful manipulation when developing agent code and the "brain" of an agent.

To support such transactions relating to content languages and ontologies JADE is designed to perform the semantic checks to ensure that \( I \) is meaningful and that it complies with the rules of the ontology set in place. The conversion and operation checks described above are carried out by a content manager object, where the
ContentManager class provides all the methods to transform Java objects into Strings or Bytes and to insert them into the content slot of an ACL messages and visa versa. The content manager provides a convenient interface to access the conversion functionality, but actually just delegates the conversion and check operations to an ontology and a content language codec. More specifically the ontology validates the information to be converted from the semantic point of view while the codec performs the translation into strings according to the syntactic rules relating to the content language chosen.

Within JADE, FIPAs agent language specification dictates all messages to have a semantic conforming to its preformative (i.e. type of action taken), where the content reference model discerns between:

- **Predicates** – which say something about the status of the world and can either be true or false, and can be meaningfully used for instance in the content of an INFORM or QUERY-IF message, while it would make no sense if used as he content of a REQUEST message
- **Concepts** – which are expressions that indicates with a complex structure that can be defined in terms of slots e.g. Person :name John :age 33), where concepts typically make no sense if used directly as the content of an ACL message, where typically they are used inside predicates and other concepts.
- **Agent Actions** – which are a special type of concept that indicate actions that can be preformed by some agents
- **Primitives** – are expressions that indicate atomic entities such as strings and integers
- **Aggregates** – are expressions indicating entities that are groups of other entities i.e. (sequence (Person :name John) (Person :name Bill))
- **Identifying Referential Expressions (IRE)** -are expressions that identify the entity form which a given predicate is true
- **Variables**- are expressions that are typically used in queries that indicate a generic element not known a-priori

A fully expressive content language should be able to distinguish between all the above types of elements.
<table>
<thead>
<tr>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>• AID</td>
</tr>
<tr>
<td>• BUA</td>
</tr>
<tr>
<td>• SPA</td>
</tr>
<tr>
<td>• TIA</td>
</tr>
<tr>
<td>• Service</td>
</tr>
<tr>
<td>o Functional Attributes</td>
</tr>
<tr>
<td>o Non Functional Attributes</td>
</tr>
<tr>
<td>• Price</td>
</tr>
<tr>
<td>• Rating</td>
</tr>
<tr>
<td>• Availability</td>
</tr>
<tr>
<td>• Throughput</td>
</tr>
<tr>
<td>• ScoringFunction</td>
</tr>
<tr>
<td>o PriceWeight</td>
</tr>
<tr>
<td>o RatingWeight</td>
</tr>
<tr>
<td>o AvailabilityWeight</td>
</tr>
<tr>
<td>o ThroughputWeight</td>
</tr>
<tr>
<td>o ScoringNo</td>
</tr>
<tr>
<td>• Winner</td>
</tr>
<tr>
<td>• Losers</td>
</tr>
<tr>
<td>• AgentAction</td>
</tr>
<tr>
<td>o Buy</td>
</tr>
<tr>
<td>o CFP (Call For Proposal)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Query</td>
</tr>
<tr>
<td>• Owns</td>
</tr>
<tr>
<td>• Assign</td>
</tr>
</tbody>
</table>

Table 5-2 TSE Ontology
In order to create the ontology in the TSE, the domain and scope of the ontology was defined. During this process a top down development process was initiated with the definition of the most general concepts and subsequent specialization of the concepts. For example, the top down development stated with the definition of the concepts such as the various entities operating within the system i.e. BUA, TIA, SPA and NOA, services and the scoring function involved in the negotiation. Further categorization of the service class resulted in the definition of its functional and non-functional attributes. Once the concepts and subsequent classes were defined the classification of the properties of the classes or slots were also outlined. For example in the sub class of the service concept, the non functional attributes stated were rating, availability, throughput and price, where for each of these properties the class of which it describes is defined along with the facets of the slots. The general outline of the ontology used is described below showing the concepts, agent actions and predicates defined within the domain of discourse.

As stated in Section 3.4.3 the content language used in the TSE is RDF(S) and was chosen over:

- FIPA SL – as the semantics in RDF (S) are much simpler than FIPA SL particularly at the lower layers of the semantic Web formalism stack
- XML – as when you attempt to describe various properties in one statement XML does not have enough semantics to appropriately represent this data and becomes very clumsy in its attempt to provide semantics whereas RDF is more expressive in capturing usable structures

In order to support RDF (S) with JADE, an RDF (S) codec was downloaded from the InfoLab in Stanford University which implements the Codec interface of JADE (jade.content.lang.Codec) and allows converting back and forth between arrays of byte in RDF (S) format and AbsContentElement (abstract descriptor of the message content). Using Protégé and the Ontology Bean Generator1, the TSE ontology described above in Table 5-2 can be generated into the Java files necessary for implementation in the TSE. With the beangenerator tool one can generate FIPA/JADE compliant ontologies from RDF(S), XML and Protégé projects.

1 http://protege.cim3.net/cgi-bin/wiki.pl?OntologyBeanGenerator
5.2 Web service Integration

Following the discussion provided in Section 3.7.7 and the findings made from the AgentCities projects a number of implementations have been developed to integrate Web services and agent functionality. WS2JADE [195] can dynamically translate Web service interfaces into agent ontologies and creates new gateway agents to handle Web service invocation requests from other JADE client agents at runtime. The current version of WS2JADE does not address the problems of communication and negotiation protocol mapping. The Web Services Agent Integration (WSAI) [196] project developed two main components: the Agent Gateway (WSAG) and the Agent Generator. The Agent Gateway does the actual translation from agents to Web services. The Agent Generator is a supporting tool for generating Gateway agents - entities which provide a concrete web service interface for a particular agent. However this project only addressed unidirectional communication between the agents and Web services. To overcome this limitation the Web Service Integration Gateway (WSIG) was developed by Whitestein Technology. The WSIG [197] is provided as an add-on for the JADE platform and supports bidirectional communication between agents and Web services employing a WSDL/UDDI/SOAP stack. Its components include the Gateway Registry, Service Description Translation, Stub creation, Web service invocation and agent service invocation.

The developers of the WSIG in [198] have also proposed the semantic enhancement of WSIG using OWL-S, as the common service description language. The paper discusses how the gateway allows agents and Web clients to invoke both atomic services and composition patterns consisting of both agent services and Web services represented as unified workflow patterns. This is achieved by using OWL-S to express the service description of both agent and Web services, with JADE agents engineered to facilitate the service composition process. While the architecture for supporting these transactions are discussed in [198], there has been no actual implementation of the semantic WSIG to date and development work has currently ceased in this project.

It is recognized in the TSE that a semantic gateway is required to support the discovery, invocation and monitoring of the WS-Resources owned by the various service providers operating in the TSE. For this purpose the current version of the WSIG is sufficient. Modifications of the WSIG in its current state had to be made to enable it to support
RDF(S) instead of FIPA SL as RDF(S) is the semantic language used in the TSE as discussed in Section 5.1.1. This change was necessary to allow FIPA RDF(S) service descriptions to be translated into UDDI tModels. The WSIG was also required to support more than the Request interaction protocol and was extended to support FIPA Subscribe Interaction protocol. In addition the xerces based codecs to bidirectionally translate ACL/SL0 into WSDL, tModels and SOAP according to the specific context, had to be modified in accordance to WS-Resource specifications. As a result of these modifications the WSIG was renamed to incorporate its new functionality for dealing with WS-Resources instead of Web services to WS-Resource Integration Gateway (WSRIG).
5.3 JADE Messaging and Measurement Issues

Each running instance of the JADE runtime environment is called a container as it can contain several agents. A set of active containers is called a platform, where a single special main container must always be active on the platform and all other containers register with it as soon as they start. In the platform, an important role is played by the main container where the FIPA service agents reside i.e. AMS and DF. Within the JADE system architecture two main forms of agent communication exist, as shown in Figure 5-6 - intra and inter platform communication. Intra platform communication involves agents residing in the same platform and JADE uses its internal Message Transport Protocols (MTPs) for implementing delivery services, where the message is not serialized but cloned and the new object reference is passed to the receiver. In Inter-platform communication, interaction among agents is achieved by the Agent Communication Channel (ACC), which is physically distributed across all the containers of the platform. Each container can be launched with one or more MTPs and the platform is able to internally route messages and select the best MTP. In the experiments conducted in the TSE the default MTP was used which was CORBA Internet Inter-ORB Protocol (IIOP) MTP based on standard SUN Object Request Broker (ORB) provided with the Java Development Kit (JDK). When communication occurs between agents on different containers, JADE uses Remote Method Invocation (RMI) to send the messages on the different containers.

Using the JADE messaging architecture it was possible to measure the agent processing times and Round Trip Time (RTT) from the initial communication start-up to confirmation of its completion. To measure time intervals in the created testbed the method long System.getTimeMillis() of the java.runtime class was used. Different Java Virtual Machine (JVM) implementations can provide different precision levels, with Sun SDK 1.4 for Windows providing a precision of 10ms. Most runtime simulations were conducted with two PCs connected by a 100Mbps Ethernet Local Area Network (LAN). Table 5-3 shows the configuration of hardware and software of the PCs used.
Model | Dell DX 280
Total Memory | 512 MB
Operating System | Microsoft Windows XP Professional
Operating System Version | Version 2002 Service Pack 2
Processor | Intel Pentium 4 CPU 1.80 GHz
Java | Sun SDK 1.4
JADE | 3.4
Jadex | 0.95 Beta
WSIG | 3[1].4
Apache Web Server | Tomcat 5.0
Java implementation of the Universal Description, Discovery, and Integration (UDDI) (JUDDI) | Version 0.9rc4

Table 5-3 Hardware and Software testbed configuration

During a measurement it was possible to identify three intervals, as shown in Figure 5-7. $Na(t)$ is the number of active agents and $N$ is the total amount of agents in the system. During $T1$ all agents are sequentially created and start competing with the CPU, and after a certain time (i.e. thicker line) period they start exchanging messages. In this phase two different phenomena, as shown in Figure 5-7, are experienced that tend to distort the measurements. The first distortion comes from the fact that not all agents are created during this time period which causes the average RTT to be lower (i.e. faster exchange) as there are less agents exchanging sending and exchanging messages over the ACC, while the second distortion emerges from the processing delay of the large number of agent creation competing for the CPU to actually create its instance causing the average RTT to be appear higher (i.e. lower exchange). It is unlikely that these two distortions cancel each other out and do not give an accurate view on the real RTT.
During the time interval $T_2$ all the couples are created and are ready to exchange messages, and hence during this time period all measurements should be taken. In the $T_3$ phase the measurement is again influenced by the lower number of agents competing for the CPU, and by the agent destruction time, similar to that observed in the $T_1$ phase.

In order to avoid the distortions experienced in the $T_1$ and $T_3$ time intervals the devised testbed environment consists of a number of TestAgents and a central SynchAgent. This testing methodology following the recommendations made in [199, 200] Upon startup the TestAgents inform the SynchAgent of its successful creation. When all TestAgents are created the SynchAgent sends a message to all registered TestAgents notifying them to commence a specified test procedure. The TestAgents can then register, search for, modify and deregister a number of Directory Facilitator Agent Descriptions (DFAD), where the SynchAgent synchronizes the TestAgents so they are accessing the DF or another resource/agent at the same time. This approach ensures that the measurements evaluated are taken only from the $T_m$ time interval when the system is working at full load performing the test requested by the SynchAgent.

![Figure 5-7 Number of active agents during the considered time intervals](image)

143
5.4 DF Scalability

A key element in the TSE and a common service to both B2B and B2C markets is the Directory Facilitator (DF) and its yellow pages service. As a result scalability of the DF and its performance at a high number of service descriptions was conducted. Results are averaged over 20 independent runs where various simulation scenarios were tried, with varying levels of agent’s, number of service descriptions, and operations. In the testbed setup the DF and AMS were located on the main container on one PC, while agents querying and manipulating the DF were located on a separate container on another PC.

One of the first tests conducted was to determine the RTT when registering a DF Agent Description (DFAD) to the DF. When registering the DFAD the agent needs to specify details of the service to be registered as outlined in Section 4.2.1. Figure 5-8 shows that the relationship between the RTT required for registration and the number of agents competing in the $T_m$ time period is linear and also demonstrates that registration increases for an agent with increasing numbers of service descriptions.

![Figure 5-8 Register service](image)

Figure 5-8 Register service

Figure 5-9 shows the RTT for a group search operations preformed by the DF. The difference between registration and searching is that when registering the TestAgent provides a complete description and an Agent Identifier (AID), whereas when searching the TestAgent provides a partial description with no AID. The search returns an array of
complete entries (with AIDs) whose attributes match the description. Taking the series '10 Service Descriptions' in Figure 5-9 as an example, all TestAgents have previously registered the same 10 service descriptions with the DF. After successful registration the SynchAgent notifies the TestAgent's to commence the search test with the search description containing all 10 service descriptions. When only one agent exists in the system, the DF returns the AID of this agent who matches the service description. As the number of agents increase from 10 to 50 the returned list also increases, as these TestAgent's too have registered the same services matching the search criteria. This figure shows that as the TestAgent's increase the search descriptions to be included in search the RTT also increases according to a polynomial function.

![Figure 5-9 Group Search Operation](image)

The search procedure outlined above can also be configured such that the TestAgents include 1 service description in its search, which will return all agents matching that search criteria. This test is repeated for each service previously registered with the DF. Figure 5-10 shows the results where a polynomial relationship again exists between the RTT and the number of agents with varying levels of previously registered service descriptions. On average an individual search operations can take approximately between 0.5 and 55 seconds, which is significantly greater than the group search operation shown in Figure 5-9 which varies between 0.5 to 35 seconds.
It is envisioned within the TSE that the SPA services and hence their description will change over time. As a result in the DF scalability analysis, modification of agents DFADs’ was also incorporated into the test scenarios. The approach undertaken in this test was similar to the search operation where again taking the series '10 Service Descriptions' as an example the TestAgent formatted a modification request to the DF with parameters set to include the TestAgent AID and the new DFAD for a particular service with included 10 service descriptions in its request. Figure 5-11 shows the results which demonstrate a linear relationship between the RTT and the number of agents competing with the DF at the same and also describes how increasing numbers of registered service descriptions by an agent increases the RTT. The above modify test can also be configured such that each service description is modified independently of each other. Figure 5-12 shows the results where again a linear relationship exists between the number of competing agents in the system and the RTT. When comparing the results to that presented in Figure 5-11 a similar performance is observed between group and individual search operations.

**Figure 5-10 Individual Search Operation**

<table>
<thead>
<tr>
<th>No of Service Descriptions</th>
<th>10 Service Descriptions</th>
<th>50 Service Descriptions</th>
<th>100 Service Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Service Description</td>
<td>0.43</td>
<td>1.4</td>
<td>2.62</td>
</tr>
<tr>
<td>10 Service Descriptions</td>
<td>0.43</td>
<td>5.43</td>
<td>11.36</td>
</tr>
<tr>
<td>50 Service Descriptions</td>
<td>1.4</td>
<td>11.36</td>
<td>20.8</td>
</tr>
<tr>
<td>100 Service Descriptions</td>
<td>2.62</td>
<td>35.39</td>
<td>53.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10    20    30    40    50</td>
</tr>
<tr>
<td>0     10     20     30     40</td>
</tr>
</tbody>
</table>
At the end of the configured test scenarios the agents deregistered their service descriptions from the DF and terminated their execution, the RTT analysis of which is shown in Figure 5-13.
From the above results, it is clear that the most intensive operation for the DF to deal with is the search operations. Taking the series 10 Service Descriptions as an example, group search operations takes 28/40/55/63/67% longer compared to registration for 10/20/30/40/50 agents respectively and takes 70/78/85/86/88% longer compared to modification operations on the varying number of agents. As a result of this within the B2C and B2B markets the search operations conducted by the MIA and TIA were minimized, where the MIA performs one search operation which passes the results to the TIA which uses the results to commerce its auction round. It is also possible to place a search constraint specifying the maximum number of returned results. Figure 5-14 and its corresponding data table show approx a 20/25/40/45% reduction in search time for 5 and 10 TestAgent's with 1/10/50/100 service descriptions respectively by setting their max results search constraint from 5 to 1. A similar case and reduction is observed with 10 TestAgents where a 15/29/39/40% reduction in search time is observed with setting the max results constraint from 10 to 5. The higher the number of service descriptions registered by the various TestAgents the greater the reduction in search time. A careful balance has to be achieved between search operations and search constraints, whereby setting a search constraint means potentially ignoring a service providers capability in providing a bid and subsequent service instance. It may not always be possible to ensure that this balance is meet, so as a result in the TSE a more sensible approach is adopted whereby reducing the number of search operations conducted during the automated negotiation procedures was conducted. In the TSE, the MIA provides this search
operation once, and passes the results of the query in the content of the ACL message to the TIA.

Registration of agents and their associated service description is the next most intensive operation where a relatively high overhead is placed on the system architecture with a large number of agents. In order to minimize the effect of this operation it is possible to set the *TestAgents* leasetime, which is a relative time after which the agents registration expires, to a high value in minimize the effect of this operation on the DF. Using the results gained from the analysis presented above it was possible to incorporate the results in the TSE in its design and operation.

![Figure 5-14 Search Operation with Search Constraints](image)

**Figure 5-14 Search Operation with Search Constraints**
5.5 WSRIG Performance

The Web Service Resource Integration Gateway (WSRIG) was designed as functional modules within a JADE agent platform. The WSRIG provides automatic, bidirectional operations allowing both FIPA compliant agents and Web services to be registered within the TSE. The TSE architecture diagram, incorporating the WSRIG is shown below in Figure 5-15, where the internet cloud represents the interoperable software components and their respective service providers that the SPAs and NOAs can collaborate with the aim of dynamic service composition. The WSRIG can directly interact with these software components through the Gateway agent which is implemented as a standard JADE agent in the WSRIG, with behaviors controlling access to a number of local components operating as independent threads. The gateway agent in the TSE is responsible for the following tasks:

- Receiving notification: from the agent platform DF than an agent has registered an ACL service description that it wishes to expose for invocation as an external Web service from the gateway UDDI that a Web service has registered a WSDL Service description to be exposed for invocation.
• Administering the mapping of newly registered ACL Service Description into tModels for publication via the Gateway UDDI and newly registered WSDL Service Descriptions using the semantics and ontologies set within the TSE for publication via the DF

• Receiving requests from platform agents to invoke an external Web service. These requests are in the form of FIPA Request or Subscribe messages. Responses from a Web service are returned to the agent in the form of FIPA Inform or Notify.

Performance evaluation of the WSRIG was conducted using the same methodology as outlined in Section 5.3 using TestAgents and a SynchAgent. Registration involves the TestAgent registering as normal with the DF and its service-type in its service description set to "web-service". The DF automatically forwards this request to the gateway agent in the WSIG. Two main operations are performed within the gateway agent to expose the TestAgent as a Web service: create a tModel (i.e. yellow pages) for UDDI and the creation of a business service which includes the translation activity of the ACL/RDF into WSDL. When publishing a WSDL service description, a service interface must first be published as a tModel before a service implementation is published as a business service. Figure 5-16, Figure 5-17, Figure 5-18 and Figure 5-19 describe the results for 1, 10, 50 and 100 TestAgents attempting to expose themselves as a Web service endpoint where the figures describe the 2 operations independently of each other. For 1 agent the process takes approx 2.5 seconds, while for 10 and 50 agents it takes 3 and 3.5 seconds respectively. For 100 agents the process takes approx 3.65 seconds. Once the agent has an endpoint address it can then communicate with Web services.
Figure 5-16 1 Agent Exposed as Web service Endpoint

Figure 5-17 10 Agents exposed as Web service Endpoint
Deregistration of the DFAD involves the removal of the tModel from the gateway agents UDDI. Performance analysis of the WSIG found that deregistration took approx between 0.9-1.1 seconds irrespective of the number of service descriptions of agents competing with varying inter message arrival rates in the system.

Subscribing and invoking a Web service involves the same operations within the gateway agent. Upon receipt of the request the message is parsed and the presence of an entry in the UDDI repository corresponding to the requested service is verified. If present a SOAP message is created to subscribe to or invoke the Web service, populated
with the parameters from the FIPA Subscribe/Request message and is sent to the Web service endpoint. If a response is expected a temporary endpoint is set up on the Axis service. This endpoint is removed once the appropriate response is received and has been passed to the gateway agent for return as a FIPA Agree/Inform response.

Figure 5-20 shows the processing of the gateway agent when *TestAgents* send a FIPA Request agent to the WSIG. The series ‘Extract ACL’ describes the processing time within the gateway agent to parse the FIPA message and verify an entry in the UDDI. The series shows that this operation is relatively consistent averaging around 2 ms. Translating the FIPA message to SOAP is the most intensive operation taking between 30-40 ms, while translating the Web service expected response from SOAP to FIPA takes about 10-15 ms, irrespective of the number of agents in the system. Figure 5-21 shows the total cumulative time from when the *TestAgents* send a FIPA Request/Subscribe message to when the requesting agent receives the FIPA Inform/Notify response, indicating that the operation has been executed, and includes the processing time of the gateway agent, the requesting agent, the Web service as well as the latency experienced by the network. As the number of agents accessing the message transport system and the gateway agent increases the delay the *TestAgent* experiences also increases linearly.

![Figure 5-20 WSIG Processing Request/Subscribe](image)

Figure 5-20 WSIG Processing Request/Subscribe
Figure 5-21 Round Trip Time for Request/Subscribe Procedure
5.6 Marketplace Performance Evaluation

The purpose of this section is to demonstrate the viability of the proposed architecture and to outline the B2C and B2B marketplace performance under different scenarios as highlighted in Figure 5-22, where performance evaluation utilized in total four PCs. On PC1 the main marketplace agents i.e. MIA, TIA, BBBA and NOA, were situated within the main container. On PC2 and PC3 the Service Provider Agents (SPAs) were executed in separate containers, while on PC4 the initiating agents such as Buyer User Agents (BUAs) in the B2C market or Service Provider Agents (SPAs) in the B2B market were executed.

When testing the message sequences the main container and its respective agents were initialized first. In order to optimize the agents performance in the main container they were executed on a standalone JADE environment without Jadex. The reason for this is that JADE provides single homogeneous and effective implementation of all the FIPA-Request-like interaction protocols, where the initiator sends a single message (i.e. it performs a single communicative act) within the scope of an interaction protocol in order to verify if the rational effect of the communicative act has been achieved or not. This implementation works both for 1:1 and 1:N conversation [201] and is a very effective means of initiating and controlling active communication sessions. In addition
these agents behavior resemble routing like activities and do not require Jadex and its BDI model.

Once the main container and its respective agents was running the SPAs on PC2 and PC3 in remote containers were started. The SPA agents used Jadex as their behavior was required to be more responsive in the face of changing circumstances where they were could commerce and negotiate with other service providers. In order for the SPAs to calculate the optimal quality, $q^*_i(\theta)$, attributes and price, $p^*_i(\theta)$, using the strategies provided in Section 4.1.2, OpenMapel API for Maple\(^1\) 11 [202] was used, which is a suite of Java routines that gives programs written in Java access to Maple routines, both built-in and user-defined. This API provided by Mapel 11 was used so that the results presented in the forthcoming section incorporated a processing delay which includes the delay for the agent to access its belief base over a number of beliefs such as the number of competitors that can provide the same service and its private cost parameters for producing the service, and is used by the SPAs for calculating and determining its bid using the predefined strategy. In the absence of such an API it is possible to write a set of Java methods to aid the SPA to calculate the configuration of its bid, but the OpenMapel API was used for is simplicity and eliminated the need to write such methods.

Once the SPAs were successfully registered with the required number of service descriptions they eMarketplace was ready to accept BUA\(\)SPA requests in both markets.

On PC4 for each experiment there existed 10 initiating agents (BUAs\(\)SPAs depending on the market). Upon startup these initiating agents send a message to the SynchAgent for the simulation parameters required for the experiment i.e. the number of requests to be sent by each agent, the number of service descriptions that the service providers have registered and the number of services involved in the request.

Upon receipt of the simulation information each initiating agent formulates a FIPA ACL Request and sends it to the MIA and follows the communication message sequences that were outlined in Chapter 4. Upon successful completion of the message sequences the MIA will notify the initiating agent, upon which the agent will send another request.

\[1\] Mapel is a powerful and intuitive tool for solving complex mathematical problems and creating rich, executable technical documents where the power of power of Maple can be used for performing everyday calculations, developing advanced mathematical models and creating user-friendly technical applications.
as per the parameters of the experiment. Once the agent has sent the required number of messages it sends a termination request to all agents in the eMarketplace to stop their execution and print their results to a file. All experiments were conducted with 10,000 requests being sent over the Tm time period, where each initiating agent sent 1000 requests each. In the B2C market, where the mobile user purchases a service on a per request basis the number of services in the simulation parameters is set to one. In the B2B market the number of services involved in the negotiation is varied depending on the number of atomic services required to form the composite service. The results presented in the following section will outline the B2C and B2B marketplace performance evaluation with reference to the main operations detailed in Chapter 4. These include: TIA Assignment as outlined in Section 4.3.2, Call for Proposals (CFP) summarized in Section 4.3.3, Calculate and Return Bid outlined in Section 4.3.4 and Winner Determination procedures detailed in Section 4.3.5.
5.6.1 B2C Marketplace Evaluation

The following section describes the performance of the B2C market in the TSE with varying numbers of service descriptions and service provider agents that match the BUAs service request. Within the presented scenario the SPAs are varied from 2 to 6, representing their respective services that take part in the service composition procedures.

Figure 5-23 describes the scenario where there are 2 SPAs involved in the auction round, whereas Figure 5-24 and Figure 5-25 describe the response of the system when there are 4 and 6 SPAs involved in the signalling process respectively. As expected in all diagrams TIA assignment procedures increase with enlarging number of service descriptions. SPA processing and winner determination procedures also increase where the TIA has to deal with an increasing numbers of bids in order to determine the winner of the auction round. Overall the response of the system is linear, as shown in Figure 5-26 which describes the Roundtrip Time (RTT) of the B2C market as the number of SPAs increase. Overall the B2C signalling process varies from 0.3 and 1.0 seconds.

Figure 5-23 B2C with 2SPAs
Figure 5-24 B2C 4 SPAs

Figure 5-25 B2C 6 SPAs
Figure 5- 26 Roundtrip Time for B2C Market
5.6.2 B2B Marketplace Evaluation

The following section outlines the performance of the B2B marketplace with varying number of service descriptions registered by varying numbers of service providers. The number of service providers matching the service request within the B2B market is varied from 2 to 4, representing the Irish mobile market at present where there are 4 incumbents (i.e. each service provider can provide the atomic service element involved in the process). In the future however, if it transpires that an open access policy within a mobile network exists and grassroots developers take hold and develop services, the number of incumbents could potentially increase to a much higher number. However it is unlikely, from a competitive perspective that these grassroots developers would focus on developing the same atomic service elements provided by the existing service providers, as it would be more viable for them to reuse existing services and develop services and application that distinguish themselves from their competitors. As a result of this, varying the number of service providers from 2 to 4 is assumed within the context of the following discussion to be reasonable.

In the B2B market the number of atomic services that take part in the business process is also adjusted, from 2 to 6, to evaluate the effect the performance of the system with increasing numbers of atomic components. Taking the Web service example that was used previously in Section 4.2.2, which was an MSN Messenger service which automatically translated messages from English to French and visa versa, in the 2 service scenario the service composed consists of the basic MSN Messenger and translation atomic service elements. In the 4 service example, the application may want to record the messages in a database and provide a counter to count how many times the application was accessed. As a result in the case 2 additional atomic services, a database and counter service need to be acquired as part of the service composition procedures. Finally in the 6 SPA scenario the owning application may want to extend the service to be able to translate also between English and German and hence needs a new translation service and may also decide that it wishes to charge for the service and hence may require a credit card service.

Figure 5-27 describes the performance of the B2B market in the TSE when there are two atomic services involved in the business process, and two SPAs capable of providing each atomic service. As expected as the number of registered service
descriptions increase for the SPA the TIA assignment procedure increases linearly, with
the other defined signalling procedures involved remaining relatively consistent with the
overall process taking 0.3/0.40/0.48/0.6 seconds for 10/50/100/200 service descriptions.
Figure 5-28 shows the same simulation setup except in this case there are four SPAs
with matching service descriptions for each atomic service which increases the number
of participants in each auction round. As the number of SPAs increase the TIA
assignment, SPA processing and winner determination procedures increase resulting in
the entire process taking 0.8/0.85/0.95/1.25 seconds, which is approximately a 50%
increase.
Figure 5-29 shows the performance evaluation of the B2B market signalling process when there are 4 services involved in the composite service and 2 SPAs matching each atomic service. When comparing this scenario to that described in Figure 5-27 the TIA assignment procedure remains relatively consistent. However the SPA processing increases as the SPAs have to remain a bid for two additional services with the winner determination procedure as increases by 40% at the TIA. Overall the signalling process approx doubles for 2 additional services involved in the process.

![Graph showing performance evaluation](image)

Figure 5-29 4 Services 2 SPAs

Figure 5-30 describes the results obtained for 4 services involved in the composition process with 4 SPAs with matching service descriptions that are willing to provide a bid. In this situation TIA assignment and SPA processing approx doubles compared to 2 Services with 4 SPAs as shown in Figure 5-29 with TIA assignment doubling and SPA processing increasing by 80% for 4 Services and 2 SPAs as shown in Figure 5-30.
Figure 5-30 4 Services 4 SPAs

Figure 5-31 describes the results obtained when there are 6 services involved in the composition process with 2 SPAs who are available to provide the service. A 33/66% increase is observed in both TIA assignment and SPA processing, with a 40/66% increase in winner determination procedures when comparing the results to 4 Services 2 SPAs and 2 Services 2 SPAs respectively. Figure 5-32 presents the results for 6 services with 4 SPAs with matching service descriptions. TIA assignment again doubles compared to 6 services 2 SPAs, SPA processing increases by 80% and winner determination increases by 30%.

Overall the observations made is that the RTT of the B2B marketplace is linear with respect to the number of service descriptions and the amount of services that participate in the business process, the relationship of which is shown in Figure 5-33 and Figure 5-34 for 2 and 4 SPAs respectively. The linear relationship as that presented below in Figure 5-33 and Figure 5-34 and above in Figure 5-26 is important as it means that the system is scalable for use in a real world marketplace scenario, where it can be expected that with more advanced enterprise computing platforms the performance of the eMarkets in the TSE can be improved upon. In addition to the performance aspects of the evaluation demonstrated the correctness and plausibility of the proposed system architecture, proving concrete proof that it actually works.
Figure 5-31 6 Services 2 SPAs

Figure 5-32 6 Services 4 SPAs
Figure 5-33 RTT for B2B market for 2 SPAs

Figure 5-34 RTT for B2B market for 4 SPAs
5.7 UMTS Performance Evaluation

The purpose of this section is to demonstrate the UMTS network related latency enabling a mobile user to purchase a service on a per request basis, where the service example is a voice call. The diagram shown above in Figure 5-35 highlights the testing focus of this section which is the TSE and the network operators UMTS network. In order for the TSE to interact with the network operator’s home network, the NOA communicates with the P-CSCF in its respective UMTS home network.

The service that the mobile user purchases can be any type of service such as an application driven service delivered to the mobile user using Application Server (ASs) or Web services. More specifically the analysis was used to determine the length of time for Data Connection Setup outlined in Section 4.3.1, the NOAs check to ensure it can support the QoS required by the BUA detailed in Section 4.3.4, and MO and MT setup required to establish a connection as per Section 4.3.6, where every service regardless of its type requires some or all of these procedures i.e. the purchase of a mobile call requires PDP context activation, MO and MT setup while an Multimedia Messaging service may require only PDP context activation and MO call setup.

A UMTS Signalling Simulator, named SigSim, was used to obtain these results where SigSim is a stochastic, event-driven simulator [203, 204, 205] implemented in Visual C++. SigSim was designed to enable simulation of the transport of signalling traffic in a...
UMTS network and has the capability to construct a cell layout and a UMTS network for the coverage area. The following section will provide an overview of SigSim and the inputs and values during the simulation, as well as the results obtained as a result of this study.

### 5.7.1 SigSim Architecture and Parameters

The UMTS network is made up of Radio Access Network (RAN) and Core Network (CN) and the elements of these networks are modeled in SigSim. Mobility and traffic models are provided to drive the movement of users and the use of services. These act as triggers for a number of signalling procedures required causing transitions between Radio Resource Control (RRC) states, resulting from activity or inactivity of the user on the bearer plane. Activation of a signalling procedure, then leads to a sequence of signalling messages being transmitted between network elements. The simulation consists of four main phases, as shown in Figure 5-36, reading inputs, initialisation, execution and post-processing.

**Reading in User Inputs**: Inputs provided to the simulation consist of network configuration data, user profile data and service related parameters. Network configuration data includes data such as the number of rows, columns, and radii of cells and the geographical distribution of cells i.e. % in dense urban, urban, suburban and rural. User profile data consists of the: average velocity of mobile per
environment; PDP session arrival rate and mean holding time; and Service subscription, while service related data consists of information required to characterise the traffic models representing the service. These are then used in Packet Switched (PS) simulations only. All the simulations will be in PS mode. The information provided is service related but may include data such as average packet size, average number of packets per file or average number of files transmitted per user session. The specific values relating to this data is shown in Appendix D.

- **Initialisation**: Given the network configuration data, a cell layout and a network topology are constructed.

- **Execution**: The user profile information is required to activate a mobility model and a session model. The mobility model is used to determine the residence time of a mobile in a cell. The call and session models determine when the mobile initiates and terminates a session. These actions then drive the simulation by invoking signalling flows. For Packet Switched simulation, once a session is activated, traffic models based on service related data, are used to determine the periods of bearer activity and inactivity in a Packet Data Protocol (PDP) session.

- **Post Processing of Data include**: Procedural delays for all signalling flows in terms of total system delays, RAN and CN delays and Call Session delays for our FTP session relating to Multimedia Message Delivery.

5.7.2 SIP Performance Evaluation

Within SigSim [204, 205] signalling flows were implemented to simulate a sequence of messages transversing through a number of network entities. The start of a signalling flow is triggered by an event such as cell crossing or arrival of a call. This causes a particular network entity, i.e. the mobile to send a message to another network entity, i.e. RNC. When the RNC receives the message, it may send a response back to the mobile via Node B. Arrival times of messages at various network elements take into consideration the message size and the transmission rate across the interface where the message is sent.
When a message arrives at a network entity, it may be queued at a buffer at the network element implementation. Each processor has a particular message processing capability which is based on measurements taken from real network elements including NodeBs, RNCs, SGSN and GGSN as well as a typical server platform such as SUN Netra servers. Further information regarding this processing capability is described in Appendix D.3. Network elements are interconnected using typical transmission links based on Ethernet, STM-1 and El, where it is assumed that 10% of the available data rate is reserved for control signalling. Further information on these network interfaces is shown in Table 5-4. In the simulation it is assumed that the NOA is connected to the P-CSCF via an Ethernet connection at 10Mbit/s, where the interface has being created and defined as Gnoa.

Using this approach allows the simulation of a large scale UMTS network model where a typical UMTS network topology is analysed for a dense urban environment, where the UTRAN consists of 784 NodeBs and 4 RNCs. The Packet Switched (PS) domain consists of 2 SGSNs and 1 GGSN. With this configuration, all core-network related and mobility signalling are accounted for. In the IMS, it is assumed that there are 1 P-CSCF, 1 S-CSCF, 1 I-CSCF and 1 MGCF, where the NOA can query the status of the home network via the Gnoa interface to the P-CSCF.

The analysis presented in this study is limited to the SIP signalling message load and time delay for different services and scenarios considered. The model keeps track of the Radio Access Network delay (RAN delay) and the Core Network delay (CN delay). The total procedural delays presented include transmission times across links, queuing and

<table>
<thead>
<tr>
<th>Interface</th>
<th>Transmission Line</th>
<th>Bandwidth (Mbits/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NodeB - RNC</td>
<td>lub</td>
<td>E1</td>
</tr>
<tr>
<td>RNC - RNC</td>
<td>lur</td>
<td>STM-1</td>
</tr>
<tr>
<td>RNC - SGSN</td>
<td>lu</td>
<td>STM-1</td>
</tr>
<tr>
<td>SGSN - GGSN</td>
<td>G</td>
<td>Ethernet</td>
</tr>
<tr>
<td>GGSN - P-CSCF</td>
<td>Gi</td>
<td>Ethernet</td>
</tr>
<tr>
<td>P-CSCF - S-CSCF</td>
<td>Mw</td>
<td>-</td>
</tr>
<tr>
<td>S-CSCF - I-CSCF</td>
<td>Mw</td>
<td>STM-1</td>
</tr>
<tr>
<td>S-CSCF - HSS</td>
<td>Cx</td>
<td>STM-1</td>
</tr>
<tr>
<td>HSS - I-CSCF</td>
<td>Cx</td>
<td>-</td>
</tr>
<tr>
<td>NOA - P-CSCF</td>
<td>Gnoa</td>
<td>Ethernet</td>
</tr>
</tbody>
</table>

Table 5-4 Interface Bandwidth
processing timed at network elements. Signalling flows implemented in the simulation can be grouped into four categories: Attach/Detach Procedures, Mobility procedures, Session procedures and Timer Expiry related procedures.

### 5.7.3 Performance Analysis

In order to obtain performance characteristics for SIP-based multimedia sessions in UMTS Release 5/6 [205, 206, 14], simulations were conducted using packet voice call setup delays for mobile originated and terminated call initiation. In order to analyze the SIP based multimedia sessions and message transmission delays the assumed SIP messages and their size are shown in Table 5-5. This table details the size of the header attributes independently of the message body which was specified using Session Description Protocol (SDP) [84] and describes the SIP Messages used in Mobile Originating and Mobile Terminating call setup in accordance with 3GPP technical specification 23.228 [22].

<table>
<thead>
<tr>
<th>SIP Session Setup</th>
<th>UPLINK MO</th>
<th>DOWNLINK MT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INVITE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Header</td>
<td>507</td>
<td>634</td>
</tr>
<tr>
<td>SDP</td>
<td>229</td>
<td>226</td>
</tr>
<tr>
<td>Total</td>
<td>736</td>
<td>860</td>
</tr>
<tr>
<td><strong>UPDATE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Header</td>
<td>353</td>
<td></td>
</tr>
<tr>
<td>SDP</td>
<td>194</td>
<td>N/A</td>
</tr>
<tr>
<td>Total</td>
<td>546</td>
<td></td>
</tr>
<tr>
<td><strong>PRACK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (SDP=0)</td>
<td>370</td>
<td>364</td>
</tr>
<tr>
<td><strong>ACK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (SDP=0)</td>
<td>314</td>
<td>314</td>
</tr>
<tr>
<td><strong>183 Session Progress</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Header</td>
<td>614</td>
<td>518</td>
</tr>
<tr>
<td>SDP</td>
<td>233</td>
<td>196</td>
</tr>
<tr>
<td>Total</td>
<td>847</td>
<td>714</td>
</tr>
<tr>
<td>Total (SDP=0)</td>
<td>361</td>
<td>361</td>
</tr>
<tr>
<td><strong>180 Ringing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Header</td>
<td>363</td>
<td>363</td>
</tr>
<tr>
<td>SDP</td>
<td>194</td>
<td>194</td>
</tr>
<tr>
<td>Total</td>
<td>557</td>
<td>558</td>
</tr>
</tbody>
</table>

**Table 5-5 SIP Session Setup Message Sizes**

The results presented below used the network model as described above in Section 5.7.2 and in Appendix D.2 with a total of 210,000 users roaming the UMTS network. Table 5-6 shows the PDP activation time which can either be mobile or Network Initiated (NI) and also PDP deactivation times. These times are important as the BUA may need to initiate a PDP context to establish a connection to the TSE, terminate the PDP context
when the TSE has determined the BUAs new network operator and initiate once again a new PDP context with the chosen network operator.

<table>
<thead>
<tr>
<th>Flow ID</th>
<th>Mean Delay (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RAN</td>
</tr>
<tr>
<td>PDP Activation</td>
<td>1.57</td>
</tr>
<tr>
<td>NI PDP Activation</td>
<td>2.64</td>
</tr>
<tr>
<td>PDP Deactivation</td>
<td>0.22</td>
</tr>
</tbody>
</table>

**Table 5-6 PDP Activation/Deactivation times**

Table 5-7 describes the signalling performance of the UMTS network for the procedures outlined in Section 4.3.4 relating to the NOA determining if it can provide a UMTS bearer service to support the call and the QoS that it can provide. In order to access the impact of the low bit rate 3.4 kb/s signalling bearer used for QoS management functions the messaging procedures were also tested using a 64 kb/s data bearer with a MAC block size of 1280, as per Release 5/6 specifications [158].

<table>
<thead>
<tr>
<th>Flow ID</th>
<th>Radio Bearer</th>
<th>Mean Delay (seconds)</th>
<th>RAN</th>
<th>Core</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Calculate and Return Bid</td>
<td>3.4 kbps/s</td>
<td></td>
<td>1.11</td>
<td>0.45</td>
<td>1.56</td>
</tr>
<tr>
<td>Network Calculate and Return Bid</td>
<td>64 kbps/s</td>
<td></td>
<td>0.60</td>
<td>0.45</td>
<td>1.05</td>
</tr>
</tbody>
</table>

**Table 5-7 NOA and Network procedures for QoS Management Functions**

**COMPR ESSED SIP SESSION SETUP**

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Uplink MO</th>
<th>Downlink MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>INVITE</td>
<td>310</td>
<td>476</td>
</tr>
<tr>
<td>SanDP</td>
<td>190</td>
<td>187</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>663</td>
</tr>
<tr>
<td>UPDATE</td>
<td>98</td>
<td>N/A</td>
</tr>
<tr>
<td>SanDP</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>119</td>
<td></td>
</tr>
<tr>
<td>PRACK</td>
<td>106</td>
<td>30</td>
</tr>
<tr>
<td>Total (SDP=0)</td>
<td>91</td>
<td>43</td>
</tr>
</tbody>
</table>

**Table 5-8 Compressed SIP Session Setup Message Sizes**
Table 5-9 and Table 5-10 show Release 5/6 packet voice call setup delays using SIP signalling. These results correspond to the messaging procedures outlined in the TSE Operation chapter in Section 4.3.6. Due to setup delay particularly evident over the RAN in these sets of results it has being proposed to compress the SIP messages using SIP message compression techniques. One such compression mechanism is the Text-based Compression using Cache and Blank (TCCB) [207, 208] or other methods as proposed by the IETF Robust Header Compression (ROHC) working group. Taking the TCCB approach and implementing its algorithm reduced the SIP Message sizes for mobile originating and mobile terminating SIP call setup as shown in Table 5-8, which reduced the RAN delay by 66% with an overall reduction in session setup delay of 46% using SigSim in MO and MT SIP call setup procedures. Other approaches to help reduce delay include the co-location of network elements in the IMS to reduce transmission across external interfaces or to base IMS network elements on high performance servers with high message throughput than current reference equipment as core network delay still contributes to over 50% delay when SIP message compression using TCCB is enabled.

<table>
<thead>
<tr>
<th>Flow ID</th>
<th>TCCB</th>
<th>Mean Delay (seconds)</th>
<th>RAN</th>
<th>Core</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MO SIP Session Set-up</td>
<td>Off</td>
<td>4.38</td>
<td>1.93</td>
<td>6.31</td>
<td></td>
</tr>
<tr>
<td>MO SIP Session Set-up</td>
<td>On</td>
<td>1.49</td>
<td>1.93</td>
<td>3.42</td>
<td></td>
</tr>
<tr>
<td>Secondary PDP Context</td>
<td>N/A</td>
<td>1.18</td>
<td>0.76</td>
<td>1.94</td>
<td></td>
</tr>
</tbody>
</table>

**Table 5-9 Mobile Originated SIP call setup delay 3.4kb/s bearer**

<table>
<thead>
<tr>
<th>Flow ID</th>
<th>TCCB</th>
<th>Mean Delay (seconds)</th>
<th>RAN</th>
<th>Core</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT SIP Session Set-up</td>
<td>Off</td>
<td>4.91</td>
<td>1.56</td>
<td>6.47</td>
<td></td>
</tr>
<tr>
<td>MT SIP Session Set-up</td>
<td>On</td>
<td>1.97</td>
<td>1.64</td>
<td>3.61</td>
<td></td>
</tr>
<tr>
<td>Secondary PDP context</td>
<td>N/A</td>
<td>1.18</td>
<td>0.76</td>
<td>1.94</td>
<td></td>
</tr>
</tbody>
</table>

**Table 5-10 Mobile Terminated SIP call setup delay 3.4kb/s bearer**

When a higher bearer rate of 64 kb/s with a MAC block size of 1280 is used on the air interface considerable improvements are observed, as shown in Table 5-11. The contribution of the RAN delay decreases, to only 27% of the total delay. Moreover if compression is applied, the RAN delay is reduced by 64%, however the core network delay is the major contributor, compression reduces the total delay by only 17%. In the mobile terminating scenario, similar results are obtained, provided in Table 5-12. The
contribution of RAN delay is higher in the MT case (32% of the total delay) and if TCCB compression is applied, the RAN delay and the total delay decreases by 55% and 17% respectively.

<table>
<thead>
<tr>
<th>Flow ID</th>
<th>TCCB</th>
<th>Mean Delay (seconds)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RAN</td>
<td>Core</td>
</tr>
<tr>
<td>MO SIP Session Set-up</td>
<td>Off</td>
<td>0.70</td>
<td>1.90</td>
</tr>
<tr>
<td>MO SIP Session Set-up</td>
<td>On</td>
<td>0.25</td>
<td>1.90</td>
</tr>
<tr>
<td>Secondary PDP Context</td>
<td>N/A</td>
<td>1.18</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Table 5-11 Mobile Originated SIP call setup delay, 64kb/s bearer

<table>
<thead>
<tr>
<th>Flow ID</th>
<th>TCCB</th>
<th>Mean Delay (seconds)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RAN</td>
<td>Core</td>
</tr>
<tr>
<td>MT SIP Session Set-up</td>
<td>Off</td>
<td>0.75</td>
<td>1.63</td>
</tr>
<tr>
<td>MT SIP Session Set-up</td>
<td>On</td>
<td>0.34</td>
<td>1.63</td>
</tr>
<tr>
<td>Secondary PDP context</td>
<td>N/A</td>
<td>1.18</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Table 5-12 Mobile Terminated SIP call setup delay, 64kb/s bearer

The above results were required to analyze the UMTS network delay to enable the mobile user purchase a service on a per request basis. In order for the BUA to enable this feature it would first of all need to obtain an IP address from its network operator requiring PDP context activation procedures. Once the BUA has acquired an IP address it can then send its service request to the TSE which commences the B2C marketplace signalling. While the results presented in Section 5.6.1 detailed the B2C marketplace performance the analysis did not outline how long it would take the NOA to determine if it can support the QoS requested by the mobile user. Table 5-7 details this procedure and hence the performance of the B2C marketplace should reflect these results where the overall timing is modified to include this procedure. As a result Figure 5-37 shows the RTT for the B2C market with the NOA QoS Management procedures using a 64kb/s bearer.

Once the B2C marketplace determines the winner in the auction round, the BUA upon receipt of this information needs to deactivate its current PDP context and initiate a new context with the chosen network. Taking a voice service as an example, once connected to the new network, MO and MT SIP session setup procedures are used to purchase a voice call on a per request basis. Within the UMTS network the large SIP message sizes and also the number of messages required for setup procedures contribute to significant
MO and MT call setup delays. SIP Message compression techniques, such as TCCB, can be used to achieve a reduction in the signalling delay over the Radio Access Network (RAN), but is less successful in reducing core network delay due to the large number of messages and the number of network elements involved in the session setup. Co-location of server functionality and careful design and selection of SIP servers are options to further reduce the delay in the core network.

![Figure 5-37 Roundtrip Time for B2C Market including NOA QoS Management Procedures](image)

<table>
<thead>
<tr>
<th>SIGNALLING PROCEDURE</th>
<th>TIME (SECONDS)</th>
<th>OVERALL TIME (SECONDS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDP Context Activation</td>
<td>2.34</td>
<td></td>
</tr>
<tr>
<td>B2C Marketplace signalling</td>
<td>1.33 - 2.05</td>
<td></td>
</tr>
<tr>
<td>PDP Context Deactivation</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>PDP Context Activation</td>
<td>2.34</td>
<td>6.46-7.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MO AND MT SETUP</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UMTS Release 5/6 on 3.4 kb/s bearer without compression</td>
<td>12.78</td>
<td>19.24-19.96</td>
</tr>
<tr>
<td>UMTS Release 5/6 on 3.4 kb/s bearer with compression</td>
<td>7.03</td>
<td>13.49-14.21</td>
</tr>
<tr>
<td>UMTS Release 5/6 on 64 kb/s bearer without compression</td>
<td>4.98</td>
<td>11.44-12.16</td>
</tr>
<tr>
<td>UMTS Release 5/6 on 64 kb/s bearer with compression</td>
<td>4.12</td>
<td>10.58-11.3</td>
</tr>
</tbody>
</table>

Table 5-13 Overall timing for purchasing a voice call on a per request basis

176
Table 5-12 provides an overall performance analysis detailing the length of time the entire process takes to enable the BUA to purchase a voice call on a per request basis. The first section of the table outlines the PDP context activation, deactivation and varying B2C marketplace performance including the NOA QoS management procedures. Because the B2C marketplace signalling varies from 1.33 to 2.05 seconds the overall timing presented shows two sets of results, one with the B2C marketplace signalling at 1.33 seconds and the other at 2.05 seconds. What the results demonstrate is that the mobile user will have to wait approx between 10.58 to 19.96 seconds depending on the UMTS architecture, the bearer rate and the compression employed. The best scenario presented is based on UMTS Release 5/6, with a bearer rate of 64kb/s with TCCB compression employed, taking approx 10.58/11.3 seconds where 97%/91% of the delay incurred is related to UMTS network latency delay. This delay can be compared to when the mobile user would use their subscribed network to connect a call where the delay experienced would be approx 6.46/7.32 sec with TCCB on/off respectively and JAJAH were approx a 9 second delay is experienced when making a call with two mobile endpoints. Compared to this scenarios purchasing a service on a per request basis using the TSE market mechanisms would result in a 39% increase in call setup delay with UMTS and a 14/20% increase when compared to JAJAH with TCCB on/off.
5.8 Conclusion

As stated in the introduction it was the purpose of this chapter to help prove the second claim presented in this thesis. This claim refers to the novel features of the B2B and B2C markets where mobile users can purchase services on a per request basis and service providers can dynamically collaborate with each other and that these transactions can take place in a reasonable time. This evaluation is important as the mobile user would not be interested in availing of the services in the TSE if they had to incur a lengthy delay in call setup procedures. This is also true for service providers where the benefits of open access within their networks must mean that they can adapt in a flexible and responsive manner and that their transactions must also occur in a reasonable time. By performing performance analysis on the designed system architecture it was possible to prove the delay incurred for these transactions, thereby verifying that the TSE is a feasible and viable architecture for service provisioning for B3G.

Within this analysis, various constituent elements were tested to gain a greater insight into the dynamic nature of the TSE, which included performance analysis of the DF, WSRIG, B2B market, B2C market and UMTS network latency. When testing the performance of TSE the testing methodology for JADE as outlined by [199, 200] was adopted in the performance analysis of the TSE. The following section will outline some of the key findings as a result of the performance analysis conducted.

A key element of the TSE and a common service to both B2B and B2C markets is the DF and its yellow pages service. As a result scalability and performance analysis of the DF was conducted as a part of this study. Findings from this study demonstrate that the search operation procedure to the DF is the most time consuming operation varying from 0.16 to 33 seconds depending on the number of agents and their respective service descriptions. However, placing a search constraint on the number of returned results however can dramatically reduce the time taken for this operation by 40% but careful design considerations must be considered in selecting the max results to be returned. Selecting too high a value will result in decreased performance but selecting a value too low may result in service providers not even being given a chance to submit a worthwhile bid to the detriment of the BUA. As a result of this design consideration it
was decided in the TSE to minimize the number of search operations conducted by performing them once in the MIA and passing the results to the TSE in the ontology. Registration operations to the DF are also computationally intensive with times varying again from 0.2 to 3.2 seconds depending on the number of agents and the number of registered service descriptions. However setting the agents leasetime parameter to a high value will minimize the amount of re-registration procedures conducted by the agents. Modify and delete operations of agents DFADs were also tested with modification procedures taking between 0.08 and 1.5 seconds and the deregistration operation taking between 0.07 to 0.90 seconds depending on the number of agents and the number of registered service descriptions.

Performance evaluation of the WSRIG and the operations related to exposing the agent as a Web service endpoint and subscribing to and invoking a Web service was also conducted. Exposing an agent as a Web service endpoint with a varying number of agents competing for the gateway agent resulted in timings varying from 1.35 to 3.53 seconds. Deregistration of an agents DFAD from the DF and UDDI was fairly consistent irrespective of the number of agents and took between 0.9 and 1.1 seconds. Subscribing to and invoking a Web service resulted in the same procedures in the WSRIG being invoked and took approx between 0.40 and 3.30 seconds. Adopting a centralized approach to the integration problem between agents and Web services does result in a single point of failure within the designed architecture. However as agents are distributed entities in nature it would be possible to distribute the gateway agent’s functionality across multiple hosts located in geographically distributed locations, eliminating this concern and potentially increasing the performance of the WSRIG.

Signalling procedures in the B2C and B2B was also evaluated during the scope of this study. When testing the B2C marketplace the number of SPAs matching the BUAs service request was adjusted from 2 to 6 to determine the effect on the different signalling procedures. In addition the number of service descriptions that each of these SPAs have registered was also varied. Results demonstrate, as shown in Figure 5-26, a linear relationship between the number of SPAs bidding in the action round and the RTT to enable the BUA to purchase a service on a per request basis. Overall the process takes between 0.28 to 1.0 second. In the B2B market the number of atomic services that take part in the business process was also adjusted, in addition to the number of SPAs
that match the composition request and the number of registered service descriptions. A linear relationship is again demonstrated between the number of services to be composed in the composite service and the RTT as shown in Figure 5-33 and Figure 5-34, where the signalling procedures range from 0.38 to 1.5 seconds for 2 SPAs and 0.75 to 3.4 seconds for 4 SPAs.

A computer simulation-based performance evaluation revealed an interesting insight into the performance of the SIP message flows for voice services. These simulations were required to analyze the UMTS network delay to enable the mobile user purchase a service on a per request basis, where Table 5-13 describes the overall results obtained. As outlined previously in Section 5.7.3 the performance of the system depends on the UMTS architecture, the bearer rate and the compression employed, where the best scenario presented is based on UMTS Release 5/6, with a bearer rate of 64kb/s with TCCB compression employed, taking approx 10.58/11.3 seconds.

In summary through quantitative analysis of the signalling procedures conducted in the TSE it is possible to verify that the operations in the:

- B2C market takes place in a reasonable time, where with a bearer rate of 64kb/s with TCCB compression employed takes 10.58-11.3 seconds, which when compared to when a mobile user using their own subscribed network to connect a call is a 39% increase in call setup and when compared to Web activated telephony such as JAJAH results in a 14/20% increase with TCCB on/off. This increase is observed to be reasonable considering the fact that the mobile user can exert their bargaining power and are endowed with the flexibility of incorporating their price and QoS constraints in the call, which is an advantage over the other approaches.

- B2B market also takes place in a reasonable time, where the focus of the presented results was on plan selection based on the service providers QoS requirements. Key findings demonstrated the service providers do become more flexible and adaptable as they are able to select service atomic elements varying in time from 1 to 4 seconds depending on the number of service providers and their respective service descriptions.

The performance analysis thereby conducted as part of this work demonstrates that the TSE is a feasible and viable architecture for service provisioning in B3G.
Chapter 6
Conclusions and Future Work

This chapter provides a summary of the work presented in this thesis and the contribution of the Telecommunication Service Exchange (TSE) has toward service provisioning for next generation networks. Directions for future work in which the present work can be extended are also outlined.
6.1 Conclusion

As detailed in Chapter 1, the field of telecommunications over the past decade has being subject to a continuous evolution as a result of: deregulation; increased competition; and technological progress. Deregulation in the mobile sector has resulted in increased competition where new players in the form of MVNOs have added to the choices that mobile users have for business and personal communication. However much work has to be done - where a strategic review of deregulation conducted by Ofcom outlines the tactical direction its activities in the future. Key aspects of this report [3] detailed the need of mobile users to have a broader range of mobile services other than simple voice and also to have greater flexibility in switching between their service providers. The issues relating to these deregulation activities are further outlined below:

- Mobile users want a broader range of services - Currently in mobile networks, voice is still the primary means of generating revenue which has allowed network and service providers to adopt a walled garden approach within their networks. However, it is envisioned that if one third of subscribers are on 3G networks by 2010 then this will mean the demand for 3G enabled data services will increase and consequently mobile users demand for a larger variety of services to be personified. In such an environment in order for service providers to maintain their competitive edge and differentiate themselves from their competitors they will have to open up access to their networks. Traditionally service and network operators have resisted this change in access policy due to competitive forces where their focus has being on protecting their existing subscriber base instead of trying to expand it. However it is far more likely that if a walled garden approach is adopted they will lose their existing subscribing base in addition to their service differentiation capacity. Besides the obvious advantage of service differentiation, openness in telecommunications will also allow grass root developers to insert more ideas and diversity into the industry allowing unfretted innovation to emerge.

From a business and technical perspective in order to allow network and service providers to develop these services they must change their current large scale macro-service delivery paradigm to focus instead on many micro services. Adopting such an approach to service provisioning will allow service providers to test
subscriber test up of new services, in addition to being able of reusing existing components increasing their flexibility and adaptability in their business processes. SOA and Web services are the technology enablers for such micro services allowing them to become fully responsive to changing mobile user requirements and competition in the sector.

- **Greater flexibility in switching between service providers** - The problem related to mobile user's inflexibility of changing from one service provider to another is two fold, where the core issue around this problem being the issue of subscription contracts. Outside of a mobile users subscription contract mobile users are allowed as a result of European legislation to take their mobile number with them when they change providers making it easier from a logistical and business point of view to change suppliers. However within their subscription contracts mobile users are not able to exert their bargaining power, where they are unable to purchase services from an alternative supplier within their subscription contracts, which causes an inefficiency of competition within telecommunications.

Based on these two requirements the focus of this thesis therefore, was to create a service provisioning platform for B3G that allowed services to be developed quickly in response to mobile user's demands, in addition to addressing the key issue of mobile user's inability to exert their bargaining power inside of their subscription contracts. The contribution of this thesis was therefore the design and specification of a novel service provisioning platform, called the Telecommunication Service Exchange (TSE).

Within the TSE there were a number of goals to the research undertaken, which were outlined in Chapter 1 and are discussed further below with regards to their achievement.

Firstly, in order to develop the concrete requirements for the TSE a state of the art to current approaches to service provisioning had to be undertaken. Within UMTS networks at present, approaches to service provisioning revolve around the IMS, gateway interfaces such Parlay, Parlay-X, JAIN and OSA. However while these approaches do increase a service providers flexibility in developing new services quickly, it does not utilize some of the key enablers for dynamic service provisioning such as agents, semantic technologies and automated negotiation, which are required for the full automation of activities related to SOA and Web services. The TSE as stated in
Chapter 2 is represented as an eMarketplace, which is required to enable the goal of Adam Smith’s invisible hand. In conducting a state of the art into marketplace approaches to service provisioning it emerged that the Digital Market Place (DMP) is the only other telecommunication approach. The DMP however focuses on the sale of calls on a per call basis however, while the TSE aims to allow service providers to become more flexible and adaptable in their business process, in addition to enabling mobile users purchase services on a per request basis. The distinction between call and service is important as the TSE through its choice of signalling and communication protocol (i.e. SIP) is session based and hence gives service providers the freedom of offer innovate services that let subscribers add users and media at will, whereas the DMP cannot do this.

In addition the DMP does not support dynamic service composition procedures enabling service providers to dynamically collaborate with each other based on their price and QoS requirements. The TSE on the other hand focuses particularly in this domain addressing the problem of component selection for participation in a business process. Work within this domain from Zeng et al [70], Wang et al [74], CIM [75] and eFlow [76] to name a few try to resolve this problem by using techniques such as task level selection, integer programming and adaptive strategies. Priest [81] and Esmaeilsabzali [82] propose the use of auctions, with Priest [81] proposing the English auction protocol and Esmaeilsabzali [82] using a multi-attribute auction protocol which is the closest work to that presented in this thesis. However Esmaeilsabzali focuses on the strategies and not on the enabling framework to allow such transactions.

Through qualitative analysis of current service provisioning goals, it was possible to achieve the first goal of the research undertaken and determine that there currently does not exist a service provisioning platform for B3G that incorporates the two aforementioned requirements. Based on the findings of this chapter it was then possible to design a service provisioning platform which overcomes the limitations of existing approaches which is the second goal of the research presented in this thesis.

Chapter 3 of this thesis therefore focuses on the design elements of the TSE using concepts and technologies which are suitable to enable competition and allow service providers to become more flexible and adaptable in their business processes. Section 3.1 of this chapter outlines the key features of the TSE specifying the principle.
requirements that the TSE must satisfy in order to enable its full potential. Using these features a number of enabling technologies such as:

- **SOA and Web services** – which enables service and network providers to become more flexible and adaptable in their business processes allowing them to rapidly respond to changing mobile user service requirements. Various research report from Forrestor and IBM were cited in Section 3.2 to support the above claim;

- **Stateful Web services** – which allow service and network providers to become more flexible and adaptable to changing market and service capability flux and system irregularities such as failure and is discussed in detail in Section 3.3. Following the presented discussion, in order for service providers and network operators to manage such service instances effectively it is a stipulation within the TSE that services are represented as stateful Web services or Web Service-Resources (WS-Resources);

- **Agents** – (detailed in Section 3.4) in the TSE allow entities to become autonomous and proactive where they are able to exist independently from their owners and make their own decisions and opportunistically adopt goals and take initiative when required, essentially allowing service providers to become more flexible and responsive in their environment. Agents have also being described as the “running programs that drive Web services” and also have been used successfully in eCommerce and automated negotiation activities providing a well reasoned argument for their incorporation within the TSE;

- **and Automated Negotiation** – In order to automate the transactions within the TSE automated negotiation techniques using a multi-attribute auction protocol was selected, which was discussed in Section 3.5. The Multi-attribute auction protocol was selected as it allowed negotiation over multiple attributes other than price, and has the advantages of auctions such as efficiency and speed of convergence. A key aspect of automated negotiation is preference elicitation, which is a mechanism that allows mobile users to specify their preferences over issues such as price and QoS, which are then used in the multi-attribute auction protocol. Section 3.6 outlines the main techniques for preference elicitation and proposes the use of AHP within the
TSE as it is relatively easy to understand and provides a rational means of approaching a decision that is very easy to explain to others.

Using these technologies and concepts the TSE architecture was presented in Section 3.7 and its corresponding negotiation model in Section 3.8.

In order to address the next goal of this thesis it was necessary to use the design components presented in Chapter 3, and demonstrate how they can be used together in the TSE to provide its contribution to service provisioning, which is presented in Chapter 4. A key aspect to Chapter 4 was to address the technical complexity issues of implementing the TSE, demonstrating to service providers how the TSE can operate within their networks. In addition this chapter outlined specifically to the regulating authorities the somewhat minor modifications of the signalling procedures within the IMS to support the TSE and its transactions. This chapter also focuses on the operation and negotiation procedures within the B2B and B2C market of the TSE and how the chosen preference elicitation technique and automated negotiation protocol is used. Through qualitative analysis of this chapter is was possible to address the first claim of this thesis by creating a service provisioning that enabled mobile users to purchase services on a per request basis and a platform where service providers can sell services where their interoperable business functionalities available over the internet can form part of business collaborations.

Chapter 5 of this thesis focused on proving both claims made within this thesis. By physically implementing the TSE architecture as presented in Chapter 3 and the designed message sequences outlined in Chapter 4, it was possible to prove that the TSE does indeed achieve its functional requirements. In addition the framework implementation allowed quantitative analysis to prove that the acquisition of a mobile services and dynamic coalition procedures take place in a reasonable time. Key findings of this chapter demonstrated with a bearer rate of 64kb/s with TCCB compression employed allow a mobile user to exert their bargaining power in approx 10.58-11.3 seconds, which when compared to when a mobile user using their own subscribed network to connect a call is a 39% increase in call setup and when compared to Web activated telephony such as JAJAH results in a 14/20% increase with TCCB on/off. Service provider coalition also takes place in a reasonable time, where the focus of the presented results was on plan selection based on the service providers QoS.
requirements. Key findings demonstrated the service providers do become more flexible and adaptable as they are able to select service atomic elements varying in time from 1 to 4 seconds depending on the number of service providers and their respective service descriptions.

As a result of the output of Chapter 4 and 5, it was then achieve the last two goals of the research conducted as discussed in Chapter 1.
6.2 Future Work

In order to enable the meaningful communication between the various services developed by different service providers, a common understanding of used terms and definitions needs to be achieved. However, if there is no global common understanding and use of terms, this understanding needs to be established between communicating parties on the fly. In order to enable this mechanism, semantic descriptions using taxonomies and ontologies need to be used. As stated in Section 3.4 the TSE used RDF(S) and developed its own ontology to enable communication within the domain of discourse. However as outlined previously RDF(S), is not expressive enough in that it does not provide enough semantics that can appropriately describe the Web service infrastructure with the amount of information it needs to fully automate the composition process. While a number of semantic markup languages have been proposed to address the limitation of RDF(S) the two most successful languages to date are OWL-S and the Web Service Modelling Ontology (WSMO). OWL-S is more mature in certain aspects compared to WSMO, including the choreography and grounding specification, however WSMO provides a more complete and conceptual model as it addresses aspects such as goals and mediators, and hence it is envisioned to be the key specification for the Semantic Web. The B2B market of the TSE specially addresses the selection of the most appropriate components to take part in the new service by looking at the service providers non-functional requirements for the new service. However the B2B market did not address the generation of the corresponding process model (potentially represented as a goal in WSMO) as a result of the selection procedures. This process model is required to define what type of service is to be constructed. These pieces of information relate to input, outputs, preconditions, conditional post-conditions and also how the process can be split in sub processes and which execution model applies them to the process components. A service provider that volunteers in providing a real implementation of a given kind of service will have to make sure that its implementation complies to the process model described within the service ontology. Further investigation into the generation of the process model in OWL-S or WSMO is warranted to enable the TSEs full potential.

Another aspect for future work relates to security, reputation and privacy issues in the TSE. Improvements in wireless communications technology over the past decade have
led to huge growth in mobile network usage. Over two billion people worldwide own and use mobile phones and the market is continuing to grow [209]. The near-ubiquity of cellular communications has opened a vast market for network services, which allows users to use handheld devices in new ways. Another trend in the last decade is the increase in threats to privacy. Demographic data has become a highly valued commodity, and personal information is now being traded in ways previously unimagined. Whole industries are dedicated to collecting, analysing, and selling sensitive data that individuals once viewed as private. In addition, privacy-invading technologies make the task to keep personal information private an ever more difficult task. Within the TSE the specific security threats must be identified as mobile users will no longer have one constant service provider, where the mobile user may interact with many different service and network providers, often in a fleeting nature. Once these specific security threats are identified measures and alteration of the message sequences and flow must be incorporated and forms an element of future work efforts in the TSE. Within this area work by McDiarmid & Irvine [209, 210] could be potentially evaluated and adapted to the context of the TSE.

In both markets of the TSE trust and reputation of the service providers is of pivotal importance as service providers and mobile users can collaborative or connect to another network operator which is owned by an intelligent agent in the market. One potential question raised by this approach is that the user has no long term relationship with the network operator and hence does not necessarily trust them to provide the service as contracted. The TSE already provides a system where trust is a core component in the decision making process of the TIA but needs to be extended to provide an ad hoc trust mechanism such as that proposed by Quinn et al [211] which is an architecture for discovering and selecting Web services based on trust recommendations. The trust mechanism in the TSE also needs a more concrete framework where the calculation and updating of a particular service providers rating is addressed. Further analysis within this domain would mean the provision of a reputation function that automatically alters the service provider or network operators rating depending on the achieved QoS characteristics compared to that agreed as a part of the automated negotiation procedures. Another aspect that could potentially fall under this category is the payment mechanisms that need to be employed and related efforts of digital cash [212] and W3C efforts on micro-payments providing semantic descriptions.
on how to provide all the information necessary to initialize a micro-payment and transfer this information to a wallet for processing.\(^1\)

Future work of the TSE could also involve testing the TSE by presenting it to a subset of the population. The objective of this case study would be to test the individual BUA and service providers' behavior during the dynamic runtime of the system. From a mobile user's perspective questions regarding their experience with their GUI and preference elicitation could be obtained, which could provide a useful insight into how the GUI could be enhanced further.

Further experimentation could be incorporated into a business studies or commerce course within Cork Institute of Technology (CIT), where groups of students would be equipped with skills and basic knowledge on strategies and price formation. Using these skills groups of students could then be given a SPA each, where their goal would be to implement their pricing and QoS strategies into them. The purpose of the experiment would be to find out the best strategy over the course of a few weeks, which can be determined by which group earns the highest utility or revenue. In order to provide an additive incentive within the experiment a prize could then be awarded to the winning group.

The work presented in this thesis thus offers much scope for further research in the area of semantic Web services, security and trust.

[1] http://www.w3.org/ECommerce/Micropayments/
Appendix A

Session Initiation Protocol

Session Initiation Protocol (SIP) [83] is an application layer protocol for initiating, modifying and terminating communication sessions over Internet Protocols (IP) networks and has been chosen by 3GPP as the signaling protocol of choice in the IMS of UMTS [22, 21] networks.

SIP is session-based protocol rather than call based giving network and service providers the freedom to offer innovative services that let subscribers add users and media at will. A session could be an IP telephony call, a multi-user conference that incorporates voice, video and data, instant messaging chat or a multi player online game. SIP can be used to invite participants to a scheduled or already existing session. Participants can be a person, an automated service or a physical device or handset. It can also be used to add or remove media to a session. As SIP is session based, service providers using SIP can tailor services that satisfy many end-users needs and run on a variety of devices instead of offering non-integrated services that are function and platform specific. By using the key features of SIP, service providers can focus on developing the aspects of a service that will most enrich the user's experience. The details of the session, such as the media, codec or sampling rate are not defined using SIP. Rather the body of the message contains a description of the session, encoded in some other protocol format, such as the Session Description Protocol (SDP) [84].

SIP transparently supports name mapping and redirection services, which supports personal mobility, allowing users to maintain a single externally visible identifier regardless of their current location. Borrowing from ubiquitous Internet Protocols, such as Hypertext Transfer Protocol (HTTP) and the Simple Mail Transfer Protocol (SMTP), SIP is text encoded and highly extensible and may be extended to accommodate features and services such as call control services, mobility and interoperability with existing telephone systems. While HTTP provides integration of content (text, audio, video link and other Web pages) on Web pages, SIP integrates different media into
sessions. SIP adopted the request/response paradigm of HTTP, and many HTTP message headers and codes, however there are some key differences between SIP and HTTP. Unlike HTTP, SIP is a peer-to-peer protocol. With HTTP, a Web server doesn't originate requests, whereas any SIP user agent can send a request to initiate or modify a session.

A.1 SIP Architectural components

Broadly speaking SIP networks consist of two basic components – SIP user agent client (UAC) and SIP network server.

- User Agent Server (UAS) - In the client server model, when sending requests or receiving responses, a SIP UA acts as the client in which case it is referred to as the User Agent Client (UAC). The receiving SIP UA acts as the server (receives requests and sends responses) and is referred to as the User Agent Server (UAS). UAC and UAS are logical entities that are contained in every SIP User Agent. There are a number of different UAS which are further outlined below [83]:

- SIP Proxy Server - is a key component in the SIP infrastructure. Its role as an edge routing server is similar to that of a Web proxy server. It provides routing capabilities and functions such as authentication, accounting, registration and security. The SIP proxy server is the first entity that receives all outgoing requests from a SIP UA. It routes the request traversing intermediate servers until it locates the server closest to the destination SIP User Agent, which forwards the request to the end recipient.

- Registrar - is a repository of user agent’s location information. The registrar accepts registration requests from user agents and places the information in a location database. A SIP registrar message will tell the registrar at which address the user will be available henceforth. Once the location or device changes the user agent has to send another SIP registrar message to the registrar.

- Redirect Servers - respond to SIP requests with an address where the SIP message should be redirected. It maps a destination address to one or more addresses and returns the new address list to the originator of the SIP request.
A.2 SIP Messages

SIP signaling is realized through the exchange of messages. There are two types of messages: requests and responses. Requests are sent to initiate some action and responses are sent as: replies to requests; acknowledging receipt of requests; and indicating the processing status. Requests and responses share a common message format which consists of a start-line, one or more header fields, and empty line indicating the end of the header fields and an optional message body [83].

<table>
<thead>
<tr>
<th>SIP Request</th>
<th>Description</th>
<th>RFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INVITE</td>
<td>Session Set-up</td>
<td>RFC 3261</td>
</tr>
<tr>
<td>ACK</td>
<td>Acknowledgement of final response to INVITE</td>
<td>RFC 3261</td>
</tr>
<tr>
<td>OPTIONS</td>
<td>Query of options and capabilities</td>
<td>RFC 3261</td>
</tr>
<tr>
<td>BYE</td>
<td>Session Termination</td>
<td>RFC 3261</td>
</tr>
<tr>
<td>CANCEL</td>
<td>CANCEL requests cancel pending transactions.</td>
<td>RFC 3261</td>
</tr>
<tr>
<td>REGISTER</td>
<td>Inform a server about their current location.</td>
<td>RFC 3261</td>
</tr>
<tr>
<td>INFO</td>
<td>Midcall Signaling transport</td>
<td>RFC 2976</td>
</tr>
<tr>
<td>PRACK</td>
<td>Provisional Response Acknowledgement</td>
<td>RFC 3262</td>
</tr>
<tr>
<td>COMET</td>
<td>Request to indicate that precondition were met</td>
<td>RFC 3262</td>
</tr>
<tr>
<td>REFER</td>
<td>Transfer User to a URL</td>
<td>RFC 3515</td>
</tr>
<tr>
<td>SUBSCRIBE</td>
<td>Request Notification of an event</td>
<td>RFC 3265</td>
</tr>
<tr>
<td>UNSUBSCRIBE</td>
<td>Cancel notification of an event</td>
<td>RFC 3265</td>
</tr>
<tr>
<td>NOTIFY</td>
<td>Transport of subscribed event notification</td>
<td>RFC 3265</td>
</tr>
<tr>
<td>MESSAGE</td>
<td>Transport of an instant message body</td>
<td>RFC 3428</td>
</tr>
<tr>
<td>UPDATE</td>
<td>Extension allow for the updating of session parameters peer to the final response</td>
<td>RFC 3311</td>
</tr>
<tr>
<td>REASON</td>
<td>Enables the ability to know why a SIP request was issued</td>
<td>RFC 3326</td>
</tr>
</tbody>
</table>

App A Table 1 SIP Requests

SIP have a number of requests most of which are shown in App A Table 1. The first six methods in App A Table 1 are defined in the basic SIP Specification [83], and provide the basic functionality to the protocol. The rest of the request messages are extensions and are defined in various specifications and standards. These requests add functionality to enhance the protocols operation.

SIP response messages are distinguished by a Status-line in the start-line, which consists of three fields: SIP version; status code; and reason phrase. The status code is a three digit code which represents the outcome of request processing. The range of values is between 100 to 699. The first digit indicates the class of the response. App A Table 2 outlines these status codes.
<table>
<thead>
<tr>
<th>Code Range</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1xx        | Provisional/Information response  
Provisional responses indicate that the associated request was received and being processed. Upon receipt of a provisional response, the request sender should stop retransmitting the request |
| 2xx        | Success - Success responses with Status codes in the range from 200 to 299 indicate that the request was received, understood, and successfully processed. |
| 3xx        | Redirection  
When further action such as a different location is need to complete a request, redirection messages are used to provide the new location or an alternative service that would satisfy the request |
| 4xx        | Client error – client error response status codes are sent when requests cannot be processed. The request failure should be because of bad syntax in the request message or simply because the request cannot be fulfilled by the responding server |
| 5xx        | Server error – server error response status codes are sent in cases where the request is valid but when the server is unable to fulfill the request. Server internal error (500) and Not Implement (501) are two examples of Server error response status codes |
| 6xx        | Global Failure – when a request cannot be fulfilled by any server, the global failure response status codes are returned. A UAS can return a global failure response with status code 603 to Decline a request to participate in a session |

App A Table 2 Response Codes
A.3 SIP Java Development

The Java Community Process (JCP) through the Java APIs for integrated Networks (JAIN) [29] initiative define Application Programming Interface (API) for using Java technologies to provide next generation telecommunications services. Three of these API developed under JCP and JAIN support SIP programming for call control, messaging, presence and location based services. These APIs are further outlined in App A Table 3.

<table>
<thead>
<tr>
<th>API</th>
<th>Java Platform</th>
<th>Target</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAIN SIP</td>
<td>J2SE</td>
<td>Client</td>
<td>Supports an asynchronous events and messaging model between providers and listeners. It provides standardized interfaces to stateful and stateless transactions as well as the stateful dialog models.</td>
</tr>
<tr>
<td>SIP Servlet</td>
<td>J2EE</td>
<td>Server</td>
<td>Defined in JSR 116 presents an abstract view of SIP based on the Java servlet API. SIP Servlets are Java based component applications that typically run in servlet containers on network servers.</td>
</tr>
<tr>
<td>SIP for J2ME</td>
<td>J2ME</td>
<td>Device</td>
<td>Defines a SIP interface for small devices that support J2ME platform and Mobile Information Device Profile (MIDP).</td>
</tr>
<tr>
<td>IMS Service API</td>
<td>J2SE</td>
<td>IMS AS</td>
<td>JSR 281- the, to provide developers with a single high-level interface to the widest variety of communications services allowing an easy path for convergence. P. Kessler best sums the promise of JSR 281: “Imagine that the technology skills threshold for developing new applications basically disappears and that developers can focus their creative efforts on the end-user experience”</td>
</tr>
</tbody>
</table>

App A Table 3 JCP SIP Programming API
Appendix B

Web services & Grid computing

B.1 Web services

B.1.1 Web Service Standardization

There is an evolving set of specifications being branded as the "Web Service Specifications," commonly referred to as WS-* family of specifications. Some of the WS-* specifications originated from standards organizations such as the World Wide Web Consortium (W3C), the Organization for the Advancement of Structured Information Standards (OASIS), Liberty Alliance and Web Service Interoperability Organization (WS-I), but a large majority are a result of ad hoc collaborations of vendors. These vendors include Microsoft, IBM, Sun and Oracle and they are divided into two main camps - Microsoft and IBM on one side, and Sun, Oracle and everyone else on the other. Other vendors and organizations, such as Global Grid Forum\(^1\), Sonic\(^2\) and TIBCO\(^3\), also collaborate based on their particular area of expertise.

The goal of these ad-hoc vendor collaborations is to eventually bring the specifications before a standards body, after a considerable amount of public feedback sessions and interoperability workshops have occurred. For some specifications, such as WS-Security, WS-BPEL and WS-Notification, this has already happened. The reason behind their ad hoc approach however, is that these companies feel that defining standards by committee is too slow and arduous, and that it is best to define standards offline until they are "almost baked" before finally bringing them to a standards body for "ratification" with a broader community. Microsoft and IBM have often been criticized for their approach to specification creation where it has been noticed that they "create specifications in a vacuum without input and hold them close to the vest while they develop them in their products and then release them to the standards organizations" [Ed. Julson, Web Services Group Manager, Sun]. Their ability to do this

is highlighted by that fact that if a standard does not appear on the Window’s platform, which commands upward to 95% of the PC market and does not appear in IBM’s UNIX and mainframe platforms, will anyone ever use the standard? This predominant vendor membership also creates competition between the standardization bodies themselves i.e. OASIS and W3C which adds another complexity dimension to Web service standardization.

Because of the number of standardization bodies, diverse interest groups and lack of cooperation between industrial and standardization bodies, a number of conflicting and overlapping standards have emerged over the past few years. For example the WS-Reliability Specification an output of the OASIS Web Services Reliable Messaging (WSRM) Technical Committee, overlaps with WS-ReliableMessaging of the WS-* family of specifications. To overcome these problems, Microsoft and IBM along with seven other vendors, founded the WS Interoperability Organization (WS-I) (which famously excluded Sun from the group’s board at first, but later allowed it to join). WS-I is not a standardization body per se, but it combines different Web service pieces in an installation ready package, called Web service ‘profiles’ and offers tools and guidelines for installing them. The first of these profiles, called the Basic Profile is referred to by Dan Sholler from Gartner as “the baseline for interoperable Web services. Customers should demand that all of their Web services-enabled technology be compliant with the Basic Profile, and that in turn will lay the foundation for Web services to fulfill their promise and provide technology independent interoperability” [213].

B.1.2 The Basic Profile

The Basic Profile 1.0 (BP 1.0) [214] consists of implementation guidelines recommending how a set of core Web service specifications should be used together to develop interoperable Web services, with the aim of significantly simplifying the task of implementing Web service solutions. The guidelines address technologies that cover the areas of messaging, description, discovery and security. BP 1.0, the standards of which are outlined below in App B Table 1, provides constraints and clarifications to these base specifications, along with the conventions about how to use them together, with the goal of promoting interoperability.
Within this basic profile there are three main standards Simple Object Access Protocol (SOAP) [215], Web Service Description Language (WSDL) [216] defined by W3C and the Universal Description Discovery Integration (UDDI) [217] defined by OASIS. These protocols are outlined below, and their association in the WS stack is shown in App B Figure 1.

<table>
<thead>
<tr>
<th>SOAP 1.1</th>
<th>WSDL 1.1</th>
<th>UDDI 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>XML 1.0</td>
<td>XML Schema Part 1: Structures</td>
<td>XML Schema Part 2: Data Types</td>
</tr>
<tr>
<td>RFC 2616: HyperText Transfer Protocol (HTTP)1.1</td>
<td>RFC 2818: HTTP over TLS Transport Layer Security</td>
<td>RFC 2965: HTTP State Management Mechanism</td>
</tr>
<tr>
<td>The Secure Sockets Layer Protocol Version 3.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**App B Table 1 BP 1.0 Standards**

- **Service Orchestration**
  - BPEL4WS, WSFL, XLANG, WS-CDL
- **Service Discovery/Publication**
  - UDDI
- **Service Description**
  - WSDL
- **XML-based Messaging**
  - SOAP
- **Data Encoding**
  - XML, XML Schema
- **Transport**
  - HTTP/HTTPS/SMTP

**Internet/Intranet**
- TCP/IP

**App B Figure 1 Web Service Technology Stack**

- The Simple Object Access Protocol (SOAP) [215] is a lightweight protocol intended for exchanging structured information in a decentralized, distributed environment between a service provider and a service requester. SOAP is a simple enveloping mechanism for XML payloads and defines a Remote Procedure Call (RPC) convention and a messaging convention. The SOAP standard is independent of any transport, while SOAP payloads can be carried in protocols such as HTTP and SMTP. Due to these characteristics, it does not matter what technology is used to implement the service consumer or provider, as long as it can issue XML messages.
As communications protocols and message formats are standardized in the Web community, it becomes increasingly important to be able to describe the communications in some structured way. Web Services Description Language (WSDL) [216] addresses this need by defining an XML grammar for network services as collections of communication endpoints capable of exchanging messages. WSDL specifies the operational characteristics of a Web service using an XML document and provides a notation to answer questions relating to service function, location and invocation.

Universal Description Discovery Integration (UDDI) [217] implements service discovery using a centralized model of one or more repositories containing information on multiple business entities and the services they provide. In short, a UDDI registry provides a standard based approach to locate a software service, to invoke that service, and to manage metadata relating to that service. Business and service data in the UDDI can be thought of in three categories: white pages, yellow pages and green pages. White pages contain information about a business such as its name, a set of multi-language text descriptions and contact information such as addresses, phone numbers, fax numbers, web site, etc., while yellow pages consist of business categorizations. Green pages specify how to bind to a service provider. They contain technical information about how to invoke a business service, including references to specifications for Web services and support for pointers to various files and URL based discovery mechanisms, if required. UDDI uses a nested data model of business, their services and related service binding information.

Another key specification related to Web services that provide a standardised method of adding a good deal of information to a SOAP message is WS-addressing. WS-addressing provides a way to specify information about a location other than a simple URI and to do this an EndpointReference needs to be created. Endpoint references convey the information needed to identify/reference a Web service/resource endpoint, and may be used in several different ways: endpoint references are suitable for conveying the information needed to access a Web service endpoint, but are also used to provide addresses for individual messages sent to and from Web services. To deal with this last usage case this specification defines a family of message information headers that allows uniform addressing of messages independent of underlying
transport. These message information headers, as shown in App B Figure 2, convey end-to-end message characteristics including addressing for source and destination endpoints as well as message identity. The only mandatory fields in the message information headers are the To and Action headers. Other useful headers include FaultTo and ReplyTo, which can be used to generate automatic responses back to a chosen endpoint in the event that an operation fails or is completed with a response.

```
<wsa:MessageID> xs:anyURI </wsa:MessageID>
<wsa:RelatesTo RelationshipType="...">xs:anyURI</wsa:RelatesTo>
<wsa:To>xs:anyURI</wsa:To>
<wsa:Action>xs:anyURI</wsa:Action>
<wsa:From>endpoint-reference</wsa:From>
<wsa:ReplyTo>endpoint-reference</wsa:ReplyTo>
<wsa:FaultTo>endpoint-reference</wsa:FaultTo>
```

App B Figure 2 WS-addressing message information headers

B.1.3 Event Driven Architecture (EDA)

Event-Driven Architecture (EDA) complements the approach undertaken in SOA in telecommunications. While SOA provides a request/response message exchange, EDA introduces long running asynchronous process capabilities. Building applications and systems around an EDA allows applications and systems to be constructed in a manner that facilitates more responsiveness, since event-driven systems are, by design, more normalized to unpredictable and asynchronous environments. This responsiveness allows new and existing applications/services to be assembled, reassembled, and reconfigured easily and inexpensively. EDA also promotes component/service reuse in addition to the health of the system to become more accurate and synchronized and closer to real-time changes. Event Driven techniques for adaptation of services/applications are well-known and are used in areas where changes in the environment of the service or the application trigger the behavioral adaptations, for example in the area of network management, control and agent collaborations [218, 219, 220]. EDA systems are typically based on Event-Condition-Action (ECA) [221] rules. The rules specify the actions that will be enabled by the occurrence of specified events if certain conditions are satisfied.

200
Java APIs for Intelligent Networks (JAIN) Service Logic Execution Environment (SLEE), shown in App B Figure 3 [5] is depicted by Gartner Research Services in their recent publication on EDAs [222] as “a visionary contribution to the software industry, advancing event-driven application servers (EDASs) and event-driven architecture (EDA)”. JAIN SLEE, specified in JSR-22 and JSR-240, defines a low latency, high throughput, standards based, service logic execution environment that specifies how carrier grade telecommunications services can be built, managed and executed. The JAIN SLEE architecture shown in Appendix B Figure 2, comprises of three core elements, management/framework entities and a component model. The management entities allow the whole JSLEE environment to be managed through Java Management Extension (JMX). JMX technology provides the tools for building distributed, Web-based, modular and dynamic solutions for managing and monitoring devices, applications, and service-driven networks. The various entities in the framework support the business logic implemented in distributed components, the so-called service building blocks, or SBB. The trace entity allows a centralized and single point for logging, alarms inform external management systems, timers invoke components in pre-defined intervals, and profiles provide the business logic with information and data during execution. Among them, the event router routes incoming and newly created events to previously registered SBBs and resources. The event router is more or less the heart of JSLEE’s event routing system. The architecture provided in JAIN SLEE is generic enough and hence envisioned to be to core enabler to EDA in SOA.
B.1.4 OWL-S and WSMO

OWL-S [223] (formally known as DAML-S) is an ontology for describing services, providing a standard vocabulary that can be used together with the other aspects of the OWL description language to create service descriptions. OWL-S provides an upper ontology, as shown in App B Figure 4, for services motivated by the need to provide three essential types of knowledge about a service, each characterised by the question it answers.

- *What does the service require of the user(s), or other agents, and provide for them?* The answer to this question is given in the "profile". Thus, the class SERVICE presents a SERVICEPROFILE. The service profile tells "what the service does", and gives the type of information needed by a service-seeking agent (or matchmaking agent) to determine whether the service meets its needs.

- *How does it work?* The answer to this question is given in the "model". Thus, the class SERVICE is described by a SERVICEMODEL. The service model tells "how the service works", that is, it describes what happens when the service is carried out. For nontrivial services, this description may be used by a service-seeking agent in at least four different ways: (1) to perform a more in-depth analysis of whether the service meets its needs; (2) to compose service descriptions from multiple services to perform a specific task; (3) during the course of the service enactment, to coordinate the activities of the different participants; and (4) to monitor the execution of the service.

- *How is it used?* The answer to this question is given in the "grounding". Thus, the class SERVICE supports a SERVICEGROUNDING. A service grounding specifies the details of how an agent can access a service. Typically a grounding will specify a communication protocol, message formats, and other service-specific details such as port numbers used in contacting the service.
The Web Service Modelling Ontology (WSMO) [224] aims to describe all relevant aspects related to general services which are accessible through a Web service interface with the ultimate goal of enabling the (total or partial) automation of tasks (e.g. discovery, selection, composition, mediation, execution and monitoring) involved in both intra and inter enterprise integration of Web service. WSMO has its conceptual basis in the Web Service Modelling Framework (WSMF) [225] where it is refining and extending this framework to develop a formal ontology and set of languages for core elements of Semantic Web Services. Following the aspects defined in the WSMF, WSMO identifies four top-level elements as the main concepts which have to be described in order to define Semantic Web Services. These include Ontologies, Web Services, Goals and Mediators. In WSMO Web service descriptions comprise the capabilities, interfaces and internal working of the service. Goals describe aspects related to user desires with respect to the requested functionality, where ontologies can be used to define the used domain terminology. Goals essentially model the user view in the Web service usage process, and are therefore a separate top level entity in WSMO. Mediators describe elements that handle interoperability problems between different WSMO elements, where it is envisioned that mediators will become the core concept to resolve incompatibilities on the data, process and protocol level.

There are a number of differences between OWL-S and WSMO. OWL-S does not separate what the user wants from the service providers. The Profile ontology is used as both an advertisement for the service and as a request to find a service. In WSMO, a Goal specifies what the user wants and the Web service description defines what the service provides through its capability. In OWL-S, the service model does not make a
clear distinction between choreography and orchestration. It is not based on any formal model, albeit that there are external works defining the formal semantics of OWL-S processes. In WSMO the choreography and orchestration are specified in the interface of the Web service description. To define logical expressions, OWL-S provides the choice between the Semantic Web Rule Language (SWRL), Declarations in RDF made Simple (DRS) and Knowledge Interchange Format (KIF). By leaving the choice of language to the user, OWL-S adds to the interoperability problem, rather than solving it. WSMO uses WSML which enables the combination of conceptual modelling with rules on only the expressibility level, allowing description to stay more explicitly within efficiency computable fragments. OWL-S is clearly more mature in certain aspects, including the choreography and grounding specification, however WSMO provides a more complete and conceptual model as it addresses aspects such as goals and mediators [226].

Semantic models provide agreement on the meaning and intended use of terms, and may provide formal and informal definitions of the entities, so there will be less ambiguity in the intended semantics of the provider. Semantics also make it possible to specify mappings between data exchanged through XML-based SOAP messages, which would be difficult to do with syntactic representation offered by the current standards. Current Web service standards operate at the syntactic level and lack semantic representation capabilities. In particular the current WSDL standard does not contain the semantic expressivity needed to represent the requirements and capabilities of Web Services — a requirement for addressing the vexing heterogeneity challenges that need to be addressed for achieving automated discovery, improved reuse and faster composition. As a result, the W3C has defined a mechanism to associate semantic annotations with Web services that are described using WSDL referred to as WSDL-S [227].
B.2 Grid Computing

B.2.1 Introduction

The origins of Grid computing came out of the early days of computer networks where using the “spare” CPU cycles was seen as an efficient and cost-effective way of getting the most of what was then very expensive hardware. Early application drivers were largely scientific computing [228, 229, 230] and included large scale distributed computing [231, 232], integration of large-scale data repositories [233] and tele-instrumentation [234]. Foster et al [235] defines the Grid as a middleware for “flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources – what we refer to as Virtual Organizations (VOs)”.

Both SOA and Grid technologies address similar issues in distributed computing principles and are evolving towards service orientation, open standards, collaboration and virtualization. Both are adopted in large-scale distributed systems and make use of the same standards provided by W3C and OASIS to provide more interoperability, application-to-application communication, flexibility and scalability. Grid principles focus on large scale sharing of distributed systems, while Web services are the software component necessary to provide remote programme-to-programme interaction, and offer interoperability and reusability [236]. Both paradigms address core issues such as large scale data transportation and management, high performance remote access, discovery authentication and authorization in the context of each application which they are used. The merger of these two approaches offers benefits in scalability, reuse and platform independence [237]. The following section will outline the standardization bodies and efforts responsible for this merger.
B.2.2 Standardization Bodies

<table>
<thead>
<tr>
<th>Body</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGA</td>
<td>Enterprise Grid Alliance (EGA) was launched in April 2004 by a group of leading IT vendors with a common goal to develop grid solutions specific to the enterprise to accelerate the adoption of the enterprise Grid</td>
</tr>
<tr>
<td>GGF</td>
<td>The Global Grid Forum (GGF) primary objectives are to promote and support the development, deployment and implementation of Grid technologies and applications via the creation and documentation of “best practices” – technical specifications, user experiences and implementation guidelines. The GGF has established itself as the world-wide body for defining the Grid architecture specification called Open Grid Service Architecture (OGSA) and Web Service Resource Framework (WS-RF), which are based on Web services.</td>
</tr>
<tr>
<td>Globus</td>
<td>The Globus Alliance is a community of organizations and individuals developing fundamental technologies behind the grid, which lets people share computing power, databases, instruments and other on-line tools securely across corporate, institutional and geographic boundaries without sacrificing local autonomy. Globus Alliance is an active member in the community of Grid Software developers, eScience and eBusiness projects.</td>
</tr>
<tr>
<td>W3C</td>
<td>The World Wide Web Consortium (W3C) develops interoperable technologies to lead the Web to its full potential</td>
</tr>
<tr>
<td>OASIS</td>
<td>OASIS is a not-for-profit, global consortium that drives the development, convergence and adoption of eBusiness standards</td>
</tr>
<tr>
<td>DMTF</td>
<td>Distributed Management Task Force (DMTF), developer of the Common Information Model (CIM) is the industry organization leading the development, adoption, and interoperability of management standards and initiatives for enterprise and Internet environments</td>
</tr>
<tr>
<td>SNIA</td>
<td>Storage Network Industry Association (SNIA) member are dedicated to ensuring that storage networks become complete and trusted across the IT industry</td>
</tr>
</tbody>
</table>

App B Table 2 Standardization efforts of the Grid

To enable the true vision of Grids in the enterprise, robust and ubiquitous standards must be set in place, and organizations need access to a wide range of industry standard building blocks and solutions that can be mixed and matched as needed. To date, however, there has been no standard way of looking at enterprise Grid computing. In response to this there are now several organizations that view the Grid from different angles, with different backgrounds- chartered to create standards for Grid computing or topics related to Grid computing. One of the objectives of the FP6 funded project called CoreGrid [238] is to co-ordinate the standardizations efforts undertaken by the various standardization bodies outlined below in App B Table 2.
B.2.3 Standardization Efforts

Grid standardization has moved towards the adoption of Web service technologies by developing two core standards, Open Grid Service Architecture (OGSA) [239, 240, 241], and its companion implementation standard, the Open Grid Services Infrastructure (OGSI) [242] and Web Service Resource Framework (WS-RF) [243], which essentially represent the intersection of grid computing and the Web services standards and their alignment with SOA principles.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WS-ResourceProperties</td>
<td>Describes associating stateful resources and Web services to produce WS-Resources, and how elements of publicly visible properties of a WS-Resource are, retrieved, changed and deleted</td>
</tr>
<tr>
<td>WS-ResourceLifetime</td>
<td>Allow a requester to destroy a WS-Resource either immediately or at a scheduled point in the future</td>
</tr>
<tr>
<td>WS-ResourceReferences</td>
<td>Annotate a WS-Addressing endpoint reference with information needed to retrieve a new endpoint reference when the current reference becomes invalid</td>
</tr>
<tr>
<td>WS-ServiceGroup</td>
<td>Create and use heterogeneous by-reference collections of Web services</td>
</tr>
<tr>
<td>WS-BaseFault</td>
<td>Describes a base fault type used for reporting errors</td>
</tr>
<tr>
<td>WS-Notifications family of specifications</td>
<td>Standard approaches to notification using a topic based publish and subscribe pattern</td>
</tr>
</tbody>
</table>

OGSA [239, 240] defines a standard mechanism for creating, naming and discovering persistent and transient Grid service instances; provides location transparency and multiple protocol bindings for service instances; and supports integration with underlying native platform facilities. To do this, OGSA defines a set of WSDL extensions, using extensibility elements allowed by WSDL, and conventions on the use of Web service for Grid computing. The OGSA standard essentially defines what Grid services are, what they should be capable of, and what technologies they be based on. OGSA, however, does not go into specifics of the technicalities of the specification. Instead, the aim is to help classify what is and what isn't a Grid system. Based on the Web service technology, the OGSI [242] specification defines a Grid service to be a Web service that conforms to a set of conventions (interfaces and behaviors) that define how a client interacts with a Grid service. They solve the stateless and non-transient problems of Web services by introducing a factory/instance model. Instead of having stateless services shared by many clients, a Grid service factory is used to create and
maintain multiple instances of the Grid service, each representing one resource. Typically, it is possible to have one-to-many, many-to-one, and many-to-many interactions between clients and instances of Grid services. These instances are transient in nature, as opposed to being bound to the lifetime of the Grid service container. In addition, the Grid service can also be configured to be a ‘notification source’, and certain clients to be ‘notification sinks’ (or subscribers). This means that if a change occurs in the Grid service, that change is notified to all the subscribers.

Since the OGSI specification in early 2002, the Web services world has evolved significantly. WS-Resource Framework (WS-RF) [243, 244] primary function was to integrate the recent developments in Web services in addition to addressing the criticisms of the Web service standardization committee that OGSA: (1) has too much stuff in one specification; (2) does not work well with existing Web service and XML tooling; (3) and is too object oriented. The WS-RF framework is primarily concerned with the creation, addressing, inspection and lifetime management of stateful resources. The framework provides the means to express state as stateful resources and codifies the relationship between Web services and stateful resources in terms of the implied resource pattern. The composition of a stateful resource and a Web service that participates in the implied resource pattern is termed a WS-resource. The framework describes the WS-Resource definition and describes how to make the properties of a it accessible through a Web service interface, and to manage and reason about the WS-resource’s lifetime. The WS-RF has two advantages relative to OGSI. First it better exploits existing XML standards, as well as emerging Web service standards such as WS-Addressing. Thus, the WS-RF is easier to implement within existing and emerging Web service toolkits and easier to exploit within the myriad of Web services interfaces in definition. App B Figure 5 describes the refactoring of the OGSI to WS-RF to include five normative WS-RF specifications plus WS-Notification [244].

The research and industry communities, under the guidance of GGF have contributed to the design and extension of OGSA. In addition to the definition process, the GGF and other OGSA, WS-RF contributors are developing an implementation process. Because OGSA can operate independently on any software base, implementation can and have been stemmed from current Grid systems such as Globus, Legion and Unicore, as well from new environments stimulated by OGSA's openness. However the Globus Toolkit
3.0 [245] and Legion are the only two organizations who have successful provided and OGSA compliant Grid toolkit. Globus Toolkit 4.0 features a new implementation of the WS-RF and the WS-Notification standards. The Globus Toolkit has been referred to as the "the de facto standard" by the New York Times and the "Most Promising New Technology" in the R&D Magazine [246]. The Grid Interoperability Project (GRIP) [247] aims to realise the interoperability of Globus and UNICORE and to work towards standards for interoperability in the Global Grid Forum.
B.2.4 The Semantic Grid

Grid tools and technologies, such as the middleware for sharing computational data are having significant impact on various fields of work including scientific research, but also in business. Meanwhile the Semantic Web is becoming "a place where data can be shared and processed by automated tools as well as by people.... Tomorrows programs must be able to share and process data even when these programs have been designed totally independently" [W3C Semantic Web Activity Statement]. The synergies of these ideas are striking and are further outlined in App B Table 4.

![App B Figure 5 Convergence of Web Services and OGSI/WS-RF Services](image)

**App B Table 4 The Grid vs The Semantic Web**

<table>
<thead>
<tr>
<th>The Grid</th>
<th>The Semantic Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>On demand transparently constructed multi-organized federations of distributed services</td>
<td>Automatically processable, machine understandable web</td>
</tr>
<tr>
<td>Distributed computing middleware</td>
<td>Distributed knowledge and information management</td>
</tr>
<tr>
<td>Programmatic integration, originally based on protocols and toolkits</td>
<td>Information integration, based on metadata, ontologies and reasoning</td>
</tr>
<tr>
<td>Information and compute power as a utility</td>
<td>Information and knowledge is the new utility</td>
</tr>
<tr>
<td>Application Pull – pioneers are application scientists with large scale collaboration problems, originally computationally oriented</td>
<td>Technology Push – Pioneers are primarily from the knowledge, agent and Artificial Intelligence communities</td>
</tr>
</tbody>
</table>

It was this observation that in 2001 led to the proposal of the Semantic Grid, shown in App B Figure 5, as a means to achieve the vision of a high degree of easy to use and
seamless automation to facilitate flexible collaborations and computations on a global scale, by means of machine processable knowledge both on and in the grid. The dynamic discovery, formation, and disbanding of ad-hoc Virtual Organizations (VOs) of third party resources requires that the Grid middleware be able to: use and process knowledge about the availability of services; their purpose; the way they can be combined and configured or substituted; and how they are discovered invoked and evolved. The Semantic Grid addresses these challenges in Grid computing and applications by adding meaning (ontologies, annotations and negotiation processes as studied in the Semantic Web and Software Agent paradigms) to the grid. In this way, the Semantic Grid not only provides a general semantic-based computational network infrastructure, but a rich seamless collection of intelligent, knowledge based services for enabling the management and sharing of complex resources and reasoning mechanisms. In the Semantic Grid knowledge and semantics are deployed explicitly for Grid applications, and for the development of innovative Grid infrastructures. The knowledge-oriented semantics goes hand-in-hand with the exploitation of techniques and methodologies from intelligent software agents and Web services representing various components of the VOs [248].

In recognition of the importance of Semantics in Grids, the Global Grid Forum (GGF) standards body chartered a Semantic Research Group [249] in 2003. The Forums XML-based description languages such as the Job Submission Description Language [250], the Data Format Description Language [251], WS-Agreement [252] and Oasis’ Security Assertion Markup language [253] all identify the role of semantics in Grid computing. In addition, WSMO in its working group report [254] outlined the possible benefits of employing WSMO for semantically describing the Grid properties for infrastructure, execution management, data, resource management, security, information and self-management services, while the NextGrid project [254] defined Workflow Ontology OWL-WS for Dynamic Grid Service Composition [256].

Grid research has been outlined as a key strategic objective of the FP6 IST programme, resulting in the formation of the Next Generation Grid (NGG) working group, whose report is outlined in [255]. It states that the NGG should be: transparent and reliable; open to wide user and provider communities; pervasive and ubiquitous; secure and

provide trust across multiple administrative domains; easy to use and program; based on standards for software and protocols; person centric; scalable and; easy to configure and manage. The NGG Research Report also outlines the vision for the Semantic Grid from the perspective of the end user, architectural and software vision.

The UK Engineering and Physical Sciences Research Council (EPSRC) and Department of Trade and Industry’s Core eScience [256] program also started its Semantic Grid initiative, aiming to integrate and bridge the efforts made in the Grid and Semantic Web communities. The context of this project was eScience based to enable scientists to generate, analyze, share and discuss their insights, experiences and results in an effective manner, where the underlying computing infrastructure is the Grid. The Semantic Grid’s research issues align with many aspects of the NGG by providing full support for a grid’s three recognized layers: computation and data, information and knowledge and provision of seamless, pervasive and secure resource use. Based on the research conducted within this domain, the eScience program developed a set of requirements for the Semantic Grid [256, 257]. Some of these requirements are additional, while some overlap with the NGG characteristics and they include: Resource Description, discovery and use; Resource description and enactment; Autonomic behavior; Security, Trust and Provenance; Metadata and Annotation; Information Integration; Context aware decision support; Automated VO Formation and Management; and Pervasive Computing.
Appendix C
Agents & Negotiation

C.1 Agents

C.1.1 Agent Communication Languages

According to Kone et al [258], Agent Communication Language (ACL) stems from the need to coordinate the actions of an agent with that of other agents. It can be used to share information and knowledge among agents in distributed computing environments, but also request the performance of a task. The main objective of ACL is to model a suitable framework that allows heterogeneous agents to interact and to communicate with meaningful statements that convey information about their environment or knowledge.

Appendix C Figure 1, show the progression of ACLs since the early days of agents, when there was little autonomy and each project had to invent its own ACL. The first significant project related to ACL was the Knowledge Query and Manipulation Language (KQML), proposed by the US Defense Advanced Research Projects Agency’s (DARPA) Knowledge Sharing effort in the late 1980s [259]. In the context of this project the researchers developed two main components: (1) a representation language for the contents of messages, called the Knowledge Interchange Format (KIF); and (2) a communication language KQML, which consists of a set of communication primitives, aiming to support interaction among agents in Multi-Agent Systems (MAS). KQML includes many primitives, all assertives or directives, which agents use to tell facts, ask queries and subscribe to services. A sample KQML message is (tell: sender A :receiver B :content “raining”). The semantics of KQML assumes a virtual knowledge base for each agent. Telling a fact corresponds to the sending agents attempt to extract something for the receiving agents knowledge base.

In the early 1990’s, France Telecom developed Arcol [260], which includes a smaller set of primitives than KQML. Again, the primitives are all assertives and directives, but unlike KQML they can be composed. Arcol has formal semantics based on Cohen and
Levesque’s [261] approach to speech acts, which presupposes that agents have beliefs and intentions, and can represent their uncertainty about various facts. Arcol gives performance conditions, which define when an agent may perform a specific communication. In Arcol, agent A can tell agent B something only if A believes it also and can establish that B does not believe it. Although theoretically nice, it is also seen by some as its main weakness, as it can be very difficult to determine whether the listening agent believes a fact or not and therefore whether a fact can be told to that agent (cf. [262]).

App C Figure 1 Progression of Agent Communication Languages

The most recent evolution of ACLs is the draft standard proposed by the Foundation for Intelligent Physical Agents (FIPA) [263]. This foundation is a nonprofit association whose objective consists of promoting the success of emerging agent-based technology and was officially accepted by the IEEE at its eleventh standards committee meeting in June 2005. It operates through an open international collaboration of companies and universities who are active members in the field. FIPA assigns tasks (ontologies, semantics, architectures, gateways and compliance) to technical committees, each of which has primary responsibility for producing, maintaining and updating the specifications applicable to its tasks. FIPAs Agent Communication Language (FIPA-ACL) is also based on speech act theory [194] and messages are also considered communicative acts, whose objective is to perform some action by virtue of being sent.
FIPA-ACL also defines interaction protocols, as detailed in App C Table 1 which deal with pre-agreed message exchange protocols for ACL messages.

<table>
<thead>
<tr>
<th>FIPA Identifier</th>
<th>Title of Interaction Protocol</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC00026</td>
<td>Request Interaction Protocol</td>
<td>Allows one agent to request another to perform some action</td>
</tr>
<tr>
<td>SC00027</td>
<td>Query Interaction Protocol</td>
<td>Allows one agent to request to perform some kind of action on another agent</td>
</tr>
<tr>
<td>SC00028</td>
<td>Request When Interaction Protocol</td>
<td>Allows an agent to request that the receiver perform some action at the time a given precondition becomes true</td>
</tr>
<tr>
<td>SC00029</td>
<td>Contract Net Interaction Protocol</td>
<td>One agent takes the role of manager and wishes to have some task preformed by one or more other agents and further wishes to optimize a function that characterizes the task. For a given task, any number of the participants may respond with a Proposal message</td>
</tr>
<tr>
<td>SC00031</td>
<td>English Auction Interaction Protocol</td>
<td>Auctioneer calls are expressed in Call for Proposals (cfp) acts, and are multicast to participants in the English auction. Participants propose bids in a propose act, and the auctioneer notifies winner in an accept-proposal act</td>
</tr>
<tr>
<td>SC00032</td>
<td>Dutch Auction Interaction Protocol</td>
<td>Models the Dutch auction by using a series of acts such as inform-start-of-auction, cfp, propose, accept and reject proposal</td>
</tr>
<tr>
<td>SC00033</td>
<td>Brokering Interaction Protocol</td>
<td>Is designed to support brokerage interactions in mediated systems and in multi-agent systems. A broker is an agent that offers a set of communication facilitation services to other agents using some knowledge about the requirements and capabilities of those agents</td>
</tr>
<tr>
<td>SC00034</td>
<td>Recruiting Interaction Protocol</td>
<td>Is designed to support recruiting interactions in mediated and multi-agent systems, where a recruiter is another type of broker agent</td>
</tr>
<tr>
<td>SC00035</td>
<td>Subscribe Interaction Protocol</td>
<td>Allows an agent to request a receiving agent to perform an action on subscription and subsequently when the referenced object changes</td>
</tr>
<tr>
<td>SC00036</td>
<td>Propose Interaction Protocol</td>
<td>Allows an agent to propose to receiving agents that the initiator will do the actions described in the propose communicative act when the receiving agent accepts the proposal</td>
</tr>
</tbody>
</table>

App C Table 1 FIPA ACL Interaction Protocol
C.1.2 Theories of Agency

One of the core research issues in the agent communication community involves the linkage between the semantic underpinnings of the ACL and the theory of agency that regulates and defines the agent's behavior. In order for the messages of an ACL to be formally coherent, these two theories need to be aligned.

A theory of agency is a general model that specifies what actions an agent can or should perform in various situations. It abstracts away from any particular implementation, and functions as a normative theory that is useful for analysis. Theories of agency for software agents are usually based on a small set of primitives derived from the propositional attitudes of philosophy and a set of axioms or axiom schema which defined their entailment relations. A complete theory of agency also includes accounts of the agent's general reasoning strategy and deductive model; its theory of action and causality; its account of planning and goal satisfaction; and it system of belief dynamic and so forth. An agent need not directly implement its theory of agency, but it must behave as if it did. Examples of the elements which compose a theory of agency include Moore's accounts of knowledge and action [264], Georgeff and Rao's Belief Desire Intention (BDI) architecture [265] and Cohen and Levesque's intention theories [261].

Moore was in many ways the pioneer of the use of logics for capturing aspects of agency. His main concern was the study of knowledge preconditions for actions – the question of what an agent needs to know in order to be able to perform some action. He formalized a model of ability, in a logic containing a modality for knowledge, and a dynamic logic-like apparatus for modeling action [266]. This formalism allowed for the possibility of an agent having incomplete information about how to achieve some goal, and performing actions in order to find out how to achieve it. Critiques of the formalism and attempts on how to improve it can be found at [267, 268].

One of the best known and influential contributions to the area of agent theory is due to Cohen and Levesque [261]. The formalism developed was originally used to develop a theory of intention, which the authors required as a pre-requisite for a theory of speech acts [269]. However, the logic has subsequently proved to be so useful for reasoning about agents that it has been used in an analysis and cooperation in multi-agent dialogue [270], as well as several studies in the theoretical foundations of cooperative problem
solving [271, 272]. A critique of Cohen and Levesque’s theory of intention may be found in [273].

Starting from Bratman’s work [265], Rao and Georgeff developed a logical framework for agent theory based on three primitive modalities: beliefs, desires and intentions. Their formalism is based on a branching model of time [cf. 274], in which agent beliefs (the agent’s knowledge about the world), desires (the objectives to be accomplished which defines desires), and intentions (the courses of actions currently under execution to achieve the agent’s desires), are accessible worlds and are themselves branching time structures. They are particularly concerned with the notion of realism, the question of how an agent’s belief about the future affects its desires and intentions. The intentions of an agent are subsets of the beliefs and desires, i.e. an agent moves towards some of the world states it desires to be true and believes to be possible. To be computationally tactical Rao and Georgeff also produced several simplifications to the theory, the most important one being that only beliefs are represented explicitly. Desires are reduced to events that are handled by redefined plan templates, and intentions are represented implicitly by the runtime stack of executed plans. Besides these components, the BDI model includes a plan library, namely a set of “recipes” representing the procedural knowledge of the agent, and an event queue where both events (either perceived from the environment or generated by the agent itself to notify an update of its belief base) and internal subgoals (generated by the agent itself while trying to achieve a desire) are stored.

An agent’s communicative behavior is among the behaviors regulated by a theory of agency. Because of this, the semantic theories that define the meaning of an ACL message must ultimately be linked to the entities provided by the agent’s baseline theory of agency. Current versions of KQML and FIPA-ACL handle the linkage between the semantic theory and the theory of agency by appealing to a simplified version of natural language speech act theory [261]. In this approach, agent communication is treated as a type of action that affects the world in the same way that physical acts affect the world. Precisely, message types of ACLs are considered as speech acts, which in turn are defined in terms of beliefs, desires and intentions.
C.1.3 SPA Agent Definition File

<!-- This is the Service Provider Agents ADF -->
<agent xmlns="http://jadex.sourceforge.net/jadex"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://jadex.sourceforge.net/jadex http://jadex.sourceforge.net/jadex-0.95.xsd"
      name="SPA" package="B2CMarket">

  <imports>
    <import>jadex.planlib.*</import>
    <import>jadex.adapter.fipa.*</import>
    <import>java.util.logging.*</import>
    <import>java.util.*</import>
    <import>jadex.util.collection.*</import>
    <import>jadex.runtime.*</import>
    <import>com.maplesoft.openmaple.Engine</import>
  </imports>

  <capabilities>
    <capability name="procap" file="jadex.planlib.Protocols"/>
    <capability name="dfcap" file="jadex.planlib.DF"/>
  </capabilities>

  <beliefs>
    <belief name="vectorWinnerDetermination" class="java.util.Vector"/>
    <belief name="vectorCalculateReturnBid" class="java.util.Vector"/>
    <belief name="noBUAProcessed" class="int">
      <fact>0</fact>
    </belief>
    <belief name="engine" class="com.maplesoft.openmaple.Engine"/>
    <belief name="localAID" class="AgentIdentifier"/>
    <belief name="registered" class="Boolean">
      <fact>false</fact>
    </belief>
    <belief name="noOfServices" class="int">
      <fact>10</fact>
    </belief>
    <belief name="SPARating" class="int"/>
    <beliefset name="ServiceNames" class="String"/>
    <beliefset name="NoOfCompetitors" class="int"/>
    <beliefset name="ServicePerformanceCost" class="double"/>
    <beliefset name="ServiceAvailabilityCost" class="double"/>
    <belief name="indexState" class="int"/>
    <belief name="startAgent" class="StartAgentPlan">
      <fact>new StartAgentPlan($scope.getExternalAccess())</fact>
    </belief>
    <belief name="SPName" class="String">
      <fact>"Service-Provider-Agent"</fact>
    </belief>
    <belief name="timeout" class="Long" exported="true">
      <fact>200000</fact>
    </belief>
  </beliefs>

  <!-- The filter for deciding which initial request messages are handled in this capability. -->
  <belief name="rp_filter" class="IFilter" exported="true">
    <fact>IFilter.NEVER</fact>
  </belief>
  <belief name="cnp_filter" class="IFilter" exported="true">
    <fact>IFilter.ALWAYS</fact>
  </belief>
</agent>
<goals>
  <maintaingoalref name="df_keep_registered">
    <concrete ref="dfcap.df_keep_registered"/>
  </maintaingoalref>

  <!-- Initiate a conversation using the fipa-request protocol. -->
  <achievegoal name="rp_initiate" exported="true">
    <parameter name="receiver" class="AgentIdentifier"/>
    <parameter name="content" class="Object"/>
    <parameter name="ontology" class="String" optional="true"/>
    <parameter name="language" class="String" optional="true"/>
    <parameter name="timeout" class="Long" optional="true"/>
    <parameter name="result" class="Object" direction="out"/>
  </achievegoal>

  <!-- Decide upon a requested task will be executed. -->
  <achievegoal name="rp_decide_request" exported="true">
    <parameter name="request" class="Object"/>
    <parameter name="result" class="Boolean" direction="out"/>
  </achievegoal>

  <!-- Execute the requested task. -->
  <achievegoal name="rp_execute_request" exported="true">
    <parameter name="request" class="Object"/>
    <parameter name="result" class="Object" direction="out" optional="true"/>
  </achievegoal>

  <!-- Make a cfp proposal. -->
  <achievegoal name="cnp_make_proposal" exported="true">
    <parameter name="task" class="Object"/>
    <parameter name="proposal" class="Object" direction="out" optional="true"/>
    <parameter name="proposal_info" class="Object" direction="out" optional="true"/>
  </achievegoal>

  <!-- every 30 minutes -->
  <performgoal name="DF_search" retrydelay="1800000" exclude="never"/>

  <!-- every five minutes -->
  <performgoal name="updatePerformanceAvailability" retrydelay="300000" exclude="never"/>

  <!-- every minute -->
  <performgoal name="updateServiceCost" retrydelay="60000" exclude="never"/>
  <performgoal name="updateSPARating" retrydelay="60000" exclude="never"/>

  <!-- Execute a cfp task. -->
  <achievegoal name="cnp_execute_task" exported="true">
    <parameter name="proposal" class="Object"/>
    <parameter name="proposal_info" class="Object" optional="true"/>
    <parameter name="result" class="Object" direction="out" optional="true"/>
  </achievegoal>

  <achievegoalref name="df_search">
    <concrete ref="dfcap.df_search"/>
  </achievegoalref>
  </goals>

<plans>
  <plan name="request_initiator_plan">
    <parameter name="receiver" class="AgentIdentifier"/>
    <goalmapping ref="rp_initiate.receiver"/>
  </plan>
</plans>
<goals ref="rp_initiate.ontology"/>
</parameter>
<parameter name="language" class="String">
<goals ref="rp_initiate.language"/>
</parameter>
<parameter name="timeout" class="String">
<goals ref="rp_initiate.timeout"/>
</parameter>
<parameter name="result" class="Object" direction="out">
<goals ref="rp_initiate.result"/>
</parameter>
<body new RPInitiatorPlan()</body>
<trigger>
goal ref="rp_initiate"/>
</trigger>
</plan>
<plan name="gui_plan">
<body new GUIPlan()</body>
</plan>
<plan name="cnp_receiver_plan" priority="-1">
<body new CFPReceiverPlan()</body>
<trigger>
<goal ref="cnp_cfp"/>
</trigger>
</plan>
<plan name="df_search_for_SPs">
<body new SearchDF()</body>
<trigger>
goal ref="DF_search"/>
</trigger>
</plan>
<!--plan name="update_Performance_Availability">
<body new UpdatePerfAvail()</body>
<trigger>
goal ref="updatePerformanceAvailability"/>
</trigger>
</plan-->
<plan name="update_service_cost">
<body new UpdateServiceCost()</body>
<trigger>
goal ref="updateServiceCost"/>
</trigger>
</plan>
<plan name="update_SPARating_cost">
<body new UpdateSPARating()></body>
<trigger>
goal ref="updateSPARating"/>
</trigger>
</plan>
</plans>
<events>
<messageevent name="cnp_cfp" type="fipa" exported="true">
<parameter name="performative" class="String" direction="fixed">
<value>SFipa.CFP</value>
</parameter>
<parameter name="protocol" class="String" direction="fixed">
<value>SFipa.PROTOCOL_CONTRACT_NET</value>
</parameter>
<parameter name="conversation-id" class="String">
<value>SFipa.createUniqueId($scope.getAgentName())</value>
</parameter>
<parameter name="reply-with" class="String">
<value>SFipa.createUniqueId($scope.getAgentName())</value>
</parameter>
<match>$beliefbase.cnp_filter.filter($messagemap)</match>
</messageevent>
</events>
<messageevent name="cnp_propose" type="fipa" exported="true">
<parameter name="performative" class="String" direction="fixed">
<value>SFipa.PROPOSE</value>
</parameter>
</messageevent>
220
C.2 Negotiation Analysis

Markets evolve, but they are also designed by entrepreneurs, economists, lawyers and engineers. Market design creates a meeting place for buyers and sellers and a format for transactions. Recently, economists and game theorists have begun to take a direct role and designed different kinds of market mechanisms. An area in which market design is well developed is in the study of auction based protocols. Leading economists such as Preston McAfee, John McMillan and Robert Wilson, among others, have successfully deployed game-theoretical analysis in order to design the bidding process in the case of the US Federal Communication Commission’s (FCC) spectrum auctions, as well as the design of energy markets. Game theory is an important methodology, but it is not the only technique for successful market design. Market design in practice is in its early stages and comprises of methodologies such as equilibrium analysis, game theory, mechanism design theory and experimental economics.

The basics of microeconomic theory were developed more than a century ago. Leon Walras’ equilibrium theory has been the most influential school of thought. When a collection of interconnected markets achieves a perfect balance of supply and demand with respect to the maximizing behaviors of self-interested economic agents, then the economy is in general equilibrium. Walras proposed a price-adjustment process called tatonnement in which agents respond to price signals for individual goods. In this system, the agent’s interactions are coordinated by a central auctioneer, who adjusts the general price level towards a balance, announcing a set of interim prices to elicit responses from the agents. In general, situations in which either no equilibrium exists or the resource allocation process fails to converge are called market failures. For example, the basics of equilibrium theory assume continuous, convex, and strongly monotone demand functions. When these conditions are violated the existence of competitive equilibrium is no longer guaranteed.

Classic equilibrium theory depicts the outcome, but not the activity of competing. Competition sets rights, thereby allowing an efficient allocation of resources. Much of what is interesting and important in competition is hidden in the background. The standard economic model of perfect competition is lacking in that it fails to explain where prices come. A perfectly competitive company does not pay attention to what any
of the other firms in the industry is doing. Instead it passively accepts the market price.

Game theoretic models, in contrast, view competition as a process of strategic decision making under conditions of uncertainty. Mechanism design theory differs from game theory in that game theory takes the rules of the game as given, while mechanism design theory asks about the consequences of different types of rules. It is a mathematical methodology concerned with the design of compensation and wage agreements, dispute resolution or resource allocation mechanism that guide the economic actors in decisions that determine the flow of resources. The main focus of mechanism design is in the design of institutions that satisfy certain objectives, assuming that the individuals interacting through the institution will act strategically and may hold private information that is relevant to the decision at hand. [111]

As a result of the above analysis, game theory and mechanism design are used to define the rules of the auction protocol, strategies and efficiency in the TSE, and are further outlined below.

C.2.1 Game Theory

Game theory is a set of analytical tools designed to help one understand the phenomena that we observe when decision makers interact. The basic assumptions that underlie the theory are that decision makers pursue well-defined exogenous objectives and take into account their knowledge and expectation or other decision makers’ behavior. In other words, decision makers can be said, to be rational and they reason strategically. Gibbons [121] and Osborne & Rubenstein [275] provide useful introductions into the subject.

Using game theory in the TSE, it assumes that there are a set of agents, \( i \), denoted by \( N \), where \( N = 1, 2, \ldots, n \), and \( i \in N \). The action (strategy) space of agents is represented as \( A_i \), where \( A_i \) is the set of all available actions to player \( i \), and an outcome \( a = (a_1, \ldots, a_n) \) is thus simply an action profile. In a Bayesian game, let agent \( i \)'s possible payoff function be represented by \( u_i(a_1, \ldots, a_n; t_i) \), where \( t_i \) is called player \( i \)'s type and belongs to a set of possible types (or type space) \( T_i \). Each type \( t_i \) corresponds to a different payoff functions that player \( i \) may have. Given this definition of a player’s type, saying that player \( i \) knows his or her own payoff function is equivalent to
saying that player $i$ knows her type. Likewise, saying that player $i$, may be uncertain about the payoff functions of the other players is equivalent to saying that player $i$ may be uncertain about the types of the other players, denoted by $t_{-i} = (t_1,...,t_{i-1},t_{i+1},...,t_n)$.

We use $T_{-i}$ to denote the set of all possible values of $t_{-i}$, and we use the probability distribution $p_i(t_{-i} | t_i)$ to denote player $i$'s belief about the other players' types, $t_{-i}$, given player $i$'s knowledge of his own type, $t_i$. A key concept in game theory is a strategy which is a complete contingent plan, or decision rule, that defines the action an agent will select in every distinguishable state of the world. For example, in an auction, the strategy of an agent would dictate what bid the agent should submit, given its type and the actions taken by the other agents. A pure strategy for player $i$ in a Bayesian game is a function which maps player $i$'s type into her strategy set, in that $a_i : T_i \rightarrow A_i$, so that $a_i(t_i)$ is the strategy choice of type $t_i$ of player $i$. A mixed strategy for player $i$ is $\alpha_i : T_i \rightarrow \Delta(A_i)$ so that $\alpha_i(a_i | t_i)$ is the probability assigned by $\alpha_i$ to action $a_i$ by type $t_i$ of player $i$. Each agent will have a preference over outcomes and will try to choose a strategy so that its preferred outcome occurs. These preferences are expressed in terms of utility functions where $u_i(a_1,...,a_n; t_i)$ is some real number and if agent $i$ prefers outcome $u_i(a_1,...,a_n; t_i) > u_i(b_1,...,b_n; t_i)$, then we understand that player $i$ likes outcome $a = (a_1,...,a_n)$ strictly better than outcome $b = (b_1,...,b_n)$. The goal of each agent is to maximize it utility.

Game theory is interested in finding equilibria. An equilibrium is a strategy profile which satisfies certain properties. The most well known equilibrium concept is the Nash equilibrium. A Nash equilibrium of a game $G$ in strategic form is defined as any outcome $(a_1^*,...,a_n^*)$ such that $(a_1^*,...,a_n^*) \succeq u_i(a_1,...,a_n)$ for all $a_i \in A_i$ holds for each player $i$. The set of all Nash equilibria of $G$ is denoted $N(G)$. In Bayesian Nash equilibrium every agent involved is assumed to share a common prior about the distribution of agent's types, type $F(t)$, such that for any particular game the agent profiles are distributed according to $F(t)$. In equilibrium each player or agent selects a strategy to maximize its expected utility in equilibrium with expected-utility maximizing strategies of other players, in other words each player's strategy must be a best response to the other player's strategies. That is, a Bayesian Nash Equilibrium is
simply a Nash equilibrium in a Bayesian game. A stronger equilibrium concept in game theory is the dominant strategy equilibrium. In the Dominant strategy equilibrium, the problem of forming a conjecture about the action choice of other players does not arise, because there is an optimal way of taking an action *independently* of the intended play of others. Games with dominant strategy equilibria are easy for agents to play since it is obvious what their optimal strategy is and they do not need to worry about what the other agents are doing. While many games do not have dominant strategy equilibria, in many practical implementations it is possible to carefully design games in order to guarantee that dominant strategies of the agents are such that agents are best off truthfully telling their true preferences to the auctioneer.

**C.2.2 Mechanism Design**

In a mechanism design problem one can imagine that each agent holds one of the "inputs" to a well formulated completely specified optimization problem, perhaps a constraint or an objective function coefficient, and that the system wide goal is to solve the specific instantiation of the optimization problem specified by the inputs [276].

The system wide goal in mechanisms design is defined with a social choice function, which selects the optimal consequence, given an agents type, where a social choice function \( f : T_1 \times \ldots \times T_n \rightarrow C \), chooses an consequence \( f(t) \in C \), given types \( t = (t_1, \ldots, t_n) \). A Mechanism \( M = (S_1, \ldots, S_n; g(.) ) \) defines the set of strategies \( S_i \) available to each agent, and an outcome rule \( g : S_1 \times \ldots \times S_n \rightarrow C \), such that \( g(s) \) is the consequence implemented by the mechanism for strategy profile \( s = (s_1, \ldots, s_n) \). Game theory is used to analyze the consequences or outcome of a mechanism. Given mechanism \( M \) with outcome function \( g(.) \), we say that the mechanism implements a social choice function \( f(t) \), if the consequence is computed with equilibrium agent strategies is a solution to the social choice function for all possible agent preferences. The equilibrium concept can be either Nash, Bayesian-Nash, dominant – or some other solution concept, so long as the strongest one is used [277, 278].

The most important properties, according to Bichler [111], for designing a mechanism include equilibrium, efficiency, stability, incentive compatibility and speed of convergence. A mechanism is efficient, meaning that the allocation resulting from the
auction is Pareto optimal, where no agent can improve its allocation without making another agent worse off. This means that the auction is efficient if there are no further gains from trade, and that the goods are allocated to the agents who value them most highly. The solution is said to be stable if there are no subset of agents who could have done better by coming to an agreement outside the mechanism. If a mechanism is stable then it is Pareto efficient, although the reverse is not true. A direct auction is incentive compatible if honest reporting of valuations is a Nash equilibrium. Sealed bid auctions are a direct mechanism whereas English and Dutch auctions are indirect [274]. A particularly strong and strategically simple case is a mechanism where truth telling is a dominant strategy. This is a desirable feature because as an agent’s decision depends only on its local information, and it gains no advantage by expending effort to model other agents. Mechanisms that require agents to learn or estimate another’s private information do not respect privacy. Speed of convergence is another important issue in markets where transactions need to occur at a rapid rate. A good example is the Dutch flower auction. Since these auctions deal with large volumes of perishable goods, each individual transaction needs to be completed quickly.

Mechanism design theory now forms an integral part of modern economics. Its varied applications include the theory of choice, optimal tax theory and the provision of public goods. The most successful application has been with regard to auction theory. Mechanism design and auction theory share commonality in the area of optimal auction design, where principles from mechanism design are combined with auction theory to design auctions that achieve the desired optimality, although different optimality measures exist. Riley and Samuelson showed that all incentive compatible mechanisms would necessarily generate the same expected revenue. Varian [274] used mechanism design theory to derive the so-called Generalized Vickrey auction.
C.2.3 Terms and Extensions to auction protocols

In many real world situations, competition and negotiation involve many quality dimensions in addition to price. In Rothkopf and Harstad’s critical essay [119], it was pointed out that it would be useful to expand the focus, because isolated, single good auctions are not the most common and interesting ones from a practical perspective. Hence, there have been several extensions to the traditional auction paradigm in recent years.

One active field of study has been *multiple unit* and *multi-object* auctions. At multi-unit auctions, the objects for sale are assumed identical, so it matters not which unit a bidder wins but rather the aggregate number of units he wins. At multi-object auctions, the objects for sale are not identical, so it matters to a bidder which specific objects he wins. Thus an example of a multi-object auction would involve the sale of an apple, orange, and a pear, while an example of a multi-unit auction would involve the sale of three identical apples. In the auction’s simplest case, the bidders are allowed to buy only one unit of merchandise. In the more realistic case, such restrictions cannot be imposed. The consequence of the additional quantity dimensions is that traditional bidding strategies and auction design mechanisms should be reconsidered and adjusted. As Bapna et al [279] and Rothkopf and Harstad [119], among others have pointed out, the strong theoretical results obtained for isolated single good auctions, are not necessarily transferable to the more complicated multiple unit situation.

Another extension is the development of *combinatorial auctions*, in which bidders desire to buy or sell bundles of goods rather than one single good. For example, a seller may want to sell several kinds of related goods where many bidders may have preferences over a combination of items. After the seller receives all the bids, it will decide a non-conflicting allocation among these goods that maximizes its revenue. These sorts of auctions are involved in many situations in the real world especially the computational issues associated with winner determination and final allocation [280, 281]. For example in the German Communication Commission’s (FCC) spectrum auction, bidders placed bids on different combinations of spectrum licenses. However, combinatorial auctions are currently rare in practice. The main problems confronted in implementing these auctions are that they have computational uncertainty, in that there is no guarantee that the winning bids for such an auction can be found in a reasonable
amount of time when the number of bidders become larger, and that the auction is
cognitively complex and can lead participants to pursue perverse bidding strategies [282].

Another extension to the traditional auction paradigm that is particularly relevant to the
TSE setting are multidimensional auctions, also referred to as multi-attribute auctions.
In such a setting, there are multiple dimensions to a transaction, such as quality,
delivery time and warranty terms that all need to be incorporated through the auction
mechanism. If we have a multidimensional auction where quantity is also a relevant
variable, we refer to it as a multiple issue auction. Perhaps since
multidimensional/multiple issue auctions hold great promise for the improvement of
B2B transactions, their development has largely been practice driven.

Any important distinction to make with regards to auctions is that there exist forward or
reverse auctions. In the forward auction the seller offers a product to numerous buyers,
where the seller “controls” the market because a product is being offered that is in
demand by a number of buyers. The price offered by the buyer continues to increase
until a theoretical rational market price is met in the market. Supply and demand sets
the price. In a reverse auction, the buyer “controls” the market because the item being
offered is available from a number of sellers. The price offered by the sellers continues
to decrease until a theoretical rational market price is achieved. The basic premise of a
reverse auction is that a sufficient supply exists and seller’s profit margins are sufficient
to offer reduced prices. The reduced price will be offered because the suppliers can
instantaneously observe the prices being offered by other sellers [135, 136].

The prices that bidders pay at auction can be determined in at least two different ways.
In uniform-price auctions all bidders pay the same per-unit price, which typically equals
in the lowest bid that won some amount of the good, while in discriminatory-price
auctions each bidder pays the amount he bids for each unit he has won. The choice
between the two pricing rules has been an issue of debate since at least Friedman [283],
who argued that, under the discriminatory-pricing rule, also referred to as pay-as-bid
auctions, bidders shave their bids. Moreover, Vickrey [284] demonstrated at single-
object auctions within IPV that the extent of the shaving was greatest at high valuations.
Friedman argued further that under uniform price rule, because the bidder was less
likely to influence the price at which he traded, he would bid closer to his valuation than
at a pay-as-bid auction. Recently Ausubel et al [285] have pointed out that, even under
the uniform pricing rule, bidders have an incentive to shave their bids. However, under
uniform-pricing, instead of shaving relatively more at high values than low values,
bidders shave relatively more at low values than at high values. The reason why bidders
shave their bids on lower-value unit's relatively more is that these bids are more likely
to be pivotal in determining the traded price. Ausubel refers to this behaviour as price
reduction.
Appendix D
IMS and SigSim

D.1 IP Multimedia Subsystem (IMS)

The IMS, as shown in App D Figure 1, introduces three main logical network elements to the existing infrastructure: Call Session Control Function (CSCF), Media Gateway Control Function (MGCF), and Media Gateway (MGW). The Home Subscribe Server (HSS) is also introduced providing user profile information similar to that of the Home Location Register (HLR). The Call Session Control Function (CSCF) is a Session Initiation Protocol (SIP) server that provides/controls multimedia services for packet switched IP terminals, both mobile and fixed. The Session Control Function can take various roles as defined in [286]. The Proxy CSCF (P-CSCF) is the mobile’s first point of contact in the visited IMS network and it main function is to act as the QoS Policy Enforcement Point (PEP) in the UMTS network. The P-CSCF forwards the SIP registration messages and session establishment messages to either the S-CSCF or I-CSCF. The Serving CSCF (S-CSCF) is the hub of all signaling functions in an IMS network. In addition to session management, the S-CSCF also performs the role of a SIP Registrar within the IMS network. It also handles the session states of services for the registered endpoint. The Interrogating CSCF (I-CSCF) is the first point of contact within the home network from a visited network. Its main job is to query the HSS and find the location of the Serving CSCF. It also can perform load balancing between the S-CSCFs with the support of the HSS and can hide the specific configuration of the home network from other network operators by providing a single point of entry into the network. The IMS supports several nodes for internetworking with legacy networks. These are the MGW and the Media Gateway Control Function. The MGCF controls one of more MGWs and the connections between the PSTN bearer and the IP stream. The MGCF receives messages from the CSCF and determines what to establish within the MGW. The primary function of the MGW is to convert media from one format to another.
App D Figure 1 IMS Functional Planes
D.2 Simulation Inputs

The inputs to the ‘SigSim’ are classified in three main categories, network configuration data, user profile data and service-related data. The following tables provide a list of parameters required for a Packet Switched UMTS system simulation run used for the diverse scenarios presented in TSE Implementation and Analysis chapter. The Network configuration data is shown in App D Table 1, the user profile data is shown in App D Table 2 while the configuration service data is shown in App D Table 3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cells: (rows x columns)</td>
<td>(28, 28)</td>
<td>Cells</td>
</tr>
<tr>
<td>Number of RNC</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Number of SGSN</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Number of GGSN</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Number of CSCF</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Number of MGCF</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Number of I-CSCF</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Number of HSS</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>URA size</td>
<td>2</td>
<td>Cells</td>
</tr>
<tr>
<td>RA size</td>
<td>4</td>
<td>Cells</td>
</tr>
<tr>
<td>Cell Radius</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dense Urban</td>
<td>250</td>
<td>meters</td>
</tr>
<tr>
<td>Urban Rural</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>Suburban</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>4000</td>
<td></td>
</tr>
</tbody>
</table>

App D Table 1 Network configuration data
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subscriber Velocity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dense Urban</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>10</td>
<td>Km/h</td>
<td></td>
</tr>
<tr>
<td>Suburban</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDP session arrival</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time and duration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session arrivals are modeled by a Poisson process which means inter-arrival times are exponentially distributed. The mean session holding time considered is 168 minutes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attach option</td>
<td>1</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Mean Attach Duration</td>
<td>3.5</td>
<td>Hours</td>
<td>Motorola reference [^87]</td>
</tr>
<tr>
<td>Mean Detach Duration</td>
<td>3.0</td>
<td>Hours</td>
<td>Estimation</td>
</tr>
</tbody>
</table>

[^87]: Motorola reference
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Service types</td>
<td>3</td>
<td>N/A</td>
<td>Packet Voice, Paging IM, Session IM</td>
</tr>
<tr>
<td>Service Type: Packet Voice</td>
<td></td>
<td></td>
<td>Service Usage</td>
</tr>
<tr>
<td>Service Usage</td>
<td></td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>Ratio MO/MT sessions</td>
<td>0.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ratio Uplink/Downlink traffic</td>
<td>0.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Bit-rate requirements</td>
<td>3.4</td>
<td>Kbps</td>
<td></td>
</tr>
<tr>
<td>Delay requirement</td>
<td>N/A</td>
<td>Seconds</td>
<td></td>
</tr>
<tr>
<td>Service Type: Paging IM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Usage</td>
<td></td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>Ratio MO/MT sessions</td>
<td>0.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ratio Uplink/Downlink traffic</td>
<td>0.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Bit-rate requirements</td>
<td>64</td>
<td>Kbps</td>
<td></td>
</tr>
<tr>
<td>Delay requirement</td>
<td>N/A</td>
<td>Seconds</td>
<td></td>
</tr>
<tr>
<td>Service Type: Session IM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Usage</td>
<td></td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>Ratio MO/MT sessions</td>
<td>0.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ratio Uplink/Downlink traffic</td>
<td>0.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Bit-rate requirements</td>
<td>64</td>
<td>Kbps</td>
<td></td>
</tr>
<tr>
<td>Delay requirement</td>
<td>N/A</td>
<td>Seconds</td>
<td></td>
</tr>
</tbody>
</table>

App D Table 3 Service-related data
D.3 Network Element Implementation

Models of the different UMTS network elements have been implemented as processing units and input buffers. The processing capacity, also known as message turn-around time, and the buffer size of each network element are set according to Motorola internal references [287, 288, 289, 290] based on measurements taken from real equipments or prototypes as shown in App D Table 4. The buffers operate a first input first output (FIFO) queuing mechanism, where:

- The Node B consists of a single processing unit and input buffer; the RNC is implemented as a unit comprising four processors and four input buffers, one for each interface to which the RNC is connected (lub, lur, IuCS and IuPS). Each input buffer is a FIFO queue.

- The SGSN is based on a single FIFO queue and multiple processor cards for processing the signalling messages. The GGSN model is assumed to be similar to the SGSN model and a similar analysis is used to determine the message turn-around times. The message turn-around times used at the SGSN and GGSN are based on an analysis of the message path within the elements for each flow and delays associated with each message [288, 289, 290] and are illustrated in App D Table 5 and App D Table 6.

- The CSCF server is composed of two network elements collocated, the P-CSCF and the S-CSCF. Both elements have a processing unit and an input buffer. The I-CSCF is similarly modelled, but is collocated with the HSS.

- The MGCF and BGCF are modelled as a single processor unit and a single input buffer.
<table>
<thead>
<tr>
<th>Network Element</th>
<th>Turnaround time (ms)</th>
<th>Buffer Size (Mbits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NodeB</td>
<td>18</td>
<td>32</td>
</tr>
<tr>
<td>RNC</td>
<td>18</td>
<td>32</td>
</tr>
<tr>
<td>SGSN</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>GGSN</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>P-CSCF</td>
<td>25</td>
<td>16</td>
</tr>
<tr>
<td>S-CSCF</td>
<td>25</td>
<td>16</td>
</tr>
<tr>
<td>I-CSCF</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>MGCF</td>
<td>50</td>
<td>32</td>
</tr>
</tbody>
</table>

*App D Table 4 Implemented Network Elements assumptions*
<table>
<thead>
<tr>
<th>Signalling Flow Type</th>
<th>Turn-around time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MO PDP Context Activation</td>
<td>39</td>
</tr>
<tr>
<td>PDP Context Deactivation</td>
<td>40</td>
</tr>
<tr>
<td>MO Service Request</td>
<td>46</td>
</tr>
<tr>
<td>NI Service Request</td>
<td>39</td>
</tr>
<tr>
<td>Routing Area Update</td>
<td>19</td>
</tr>
<tr>
<td>SIP Signalling</td>
<td>50</td>
</tr>
</tbody>
</table>

**App D Table 5 SGSN Turn-around times**

<table>
<thead>
<tr>
<th>Signalling Flow Type</th>
<th>Turn-around time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MO PDP Context Activation</td>
<td>75</td>
</tr>
<tr>
<td>NI PDP Context Activation</td>
<td>59</td>
</tr>
<tr>
<td>PDP Context Deactivation</td>
<td>75</td>
</tr>
<tr>
<td>SIP Signalling</td>
<td>50</td>
</tr>
</tbody>
</table>

**App D Table 6 GGSN Turn-around times**
Bibliography


[22] 3GPP TS 23.228, *IP Multimedia Subsystem (IMS)*, Stage 2, Release 7


[26] 3GPP TS 22.127, *Service Requirement for the Open Services Access (OSA)*, Stage 1,
Release 7


241
Intelligent Systems Technologies, Mexico, 1996


S. McIlraith, T.C. Son, *Adapting Golog for composition of Semantic Web services*, in Proc. of the 8th International Conference on Knowledge Representation and Reasoning, France, April 2002


[82] S. Esmaeilsabzali, k. Larson, Service allocation for composite Web services based on Quality Attributes, in Proc. of the 7th International Conference on eCommerce Technology Workshops, Munich, Germany, 2005


[92] Oasis, *WS-BaseNotification 1.3*


[96] I. Foster, N. R. Jennings, C. Kesselman, Brain Meets Brawn: Why Grid and Agents Need Each Other, in Proc. 3rd Int. Conf. on Autonomous Agents and Multi-Agent Systems, New York, USA, 2004


[98] S. Ambroszkiewicz, Agent Based Approach to Service Description and Composition, Lecture Notes in Computer Science, Volume 2564/2003


246


configurable auction server for human and server agents, in Proc of the 2nd International conference on Autonomous Agents, pgs. 301-308, Minneapolis, USA, 1998


[128] G. Tesauro, J. Bredin, Strategic sequential bidding in auctions using dynamic programming, in Proc. in the first international joint conference on Autonomous Agents and Multi-Agent Systems, Italy, 2002


[133] R. Das, J. Hanson, J.O. Kephart, G. Tesauro, Agent-Human Interactions in the
Continuous Double Auction, in Proc. International Joint Conference of Artificial Intelligence, Seattle, USA, August 2001


249


[159] 3GPP TS 22.105, Services and Service Capabilities, Release 8


250


[169] Internet Assigned Numbers Authority IANA, [www.iana.org](http://www.iana.org)

[170] FIPA, FIPA Request Interaction Protocol, Specification number: SC00026


[175] 3GPP TS 24.228, *Signalling flows for the IP multimedia call control based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP)*, Release 6


[184] 3GPP TS29.198, Service Requirement for the Open Services Access (OSA), Release 7


Conference, Glasgow, Scotland, 2003


[208] RFC 3320, *Signalling Compression (SipComp)*, January 2003


[214] Web Services Interoperability Organization (WS-I) Basic Profile 1.0


[217] OASIS, UDDI Specification Technical Committee Draft


[219] K. Liu, L. Sun, A. Dix, M. Narasipuram, *Norm based agency for designing collaborate*


[223] OWL-S: Semantic Markup for Web Services, W3C Member Submission, November 2004


[236] C. Goble, D. De Roure, The Semantic Grid: Myth Busting and Bridge Building, in 16th European Conference on Artificial Intelligence, Valencia, Spain, 2004


[242] Global Grid Forum, Open Grid Service Infrastructure, version 1.0


[244] Globus, From Open Grid Service Infrastructure to WS-Resource Framework: Refactoring & Evolution, version 1.1


[246] Globus Project News Digest, Issue #1, November 12, 2002


[251] Global Grid Forum, *Data Format Description Language – Structural Description version 2.0*


