CAPACITY ANALYSIS AND DATA CONCENTRATION FOR SMART GRID COMMUNICATION NETWORKS AT THE POWER DISTRIBUTION LEVEL

A Dissertation by

Babak Karimi

Master of Science, Amirkabir University of Technology, 2008
Bachelor of Science, Sadjad Institute of Higher Education, 2006

Submitted to the Department of Electrical Engineering and Computer Science and the faculty of the Graduate School of Wichita State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

December 2014
CAPACITY ANALYSIS AND DATA CONCENTRATION FOR SMART GRID COMMUNICATION NETWORKS AT THE POWER DISTRIBUTION LEVEL

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Electrical Engineering.

____________________________________
Vinod Namboodiri, Committee Chair

____________________________________
Ward Jewell, Committee Member

____________________________________
Visvakumar Aravinthan, Committee Member

____________________________________
Murtuza Jadliwala, Committee Member

____________________________________
Davood Askari, Committee Member

Accepted for the College of Engineering

____________________________________
Royce Bowden, Dean

Accepted for the Graduate School

____________________________________
Abu Masud, Interim Dean
DEDICATION

To my parents
ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my advisor, Dr. Vinod Namboodiri, for his remarkable advice and support during my years in the Ph.D. program. I especially thank him for his encouragement, valuable suggestions, committed effort, and support, all of which helped me complete my research. The past few years have been a growth experience for me. I feel that being a doctoral student under his supervision has been a very special honor.

My sincere gratitude goes to my dissertation committee members, Dr. Ward Jewel, Dr. Visvakumar Aravinthan, Dr. Mutuza Jadliwala, and Dr. Davood Askari. Thank you for your time and support with my research and graduate studies at various stages.

I also thank my parents as well as my brother and sister, Rouzbeh and Panteha, for their unconditional love, encouragement, and continuous support, all of which were instrumental in this endeavor. There were times when I wondered if I could have survived graduate school without the love and encouragement of my family.

I would like to thank Parisa Nazaran, Anm Badrudoza, Vishnu Dev and Mohammad Heidari for being good friends during happy and also difficult times. We had a great time together in the Wireless Networking and Energy Systems Research Laboratory at Wichita State University.

Last, but not least, I extend my sincere acknowledgement to everyone who helped me during my graduate studies at Wichita State University. I apologize for not thanking everyone personally.
ABSTRACT

The “smart grid” generally refers to a class of digital technology that allows for two-way communication between the electric power utility and its customers, as well as sensing along the transmission and distribution lines. Smart grids offer many benefits to utilities and consumers—mostly seen as large improvements in energy efficiency on the electricity grid and in homes and offices. Little is known about how different communication architectures compare, what data carrying capacities they offer, and how to solve data collection and management problems that may arise. This dissertation specifically focuses on these challenges from the perspective of the power distribution network.

In the first part of this work, possible communications technologies for the power distribution level are suggested and compared, and a wireless mesh network architecture proposed and shown to meet the communication requirements of the power distribution system. In the second part of this dissertation, a linear chain multi-hop wireless communication architecture is proposed and shown through analysis and simulations to meet application requirements in terms of data-carrying capacity. Finally, in the last part of this dissertation, the looming issue of how to communicate and handle consumer data collected by electric utilities and manage available communication network resources is considered.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>CONTRIBUTIONS AND DISSERTATION ORGANIZATION</td>
<td>4</td>
</tr>
<tr>
<td>2.</td>
<td>COMMUNICATION OPTIONS AT DISTRIBUTION LEVEL</td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>INTRODUCTION</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>BACKGROUND AND RELATED WORK</td>
<td>8</td>
</tr>
<tr>
<td>2.3</td>
<td>CHOICE OF COMMUNICATION MEDIUM</td>
<td>11</td>
</tr>
<tr>
<td>2.4</td>
<td>IMPACT OF INTERFERENCE DUE TO TRANSMISSION LINES ON WIRELESS MEDIUM</td>
<td>13</td>
</tr>
<tr>
<td>2.5</td>
<td>TYPES OF WIRELESS COMMUNICATION ARCHITECTURES</td>
<td>14</td>
</tr>
<tr>
<td>2.6</td>
<td>WIRELESS COMMUNICATION NETWORK ARCHITECTURE FOR DISTRIBUTION SYSTEM</td>
<td>14</td>
</tr>
<tr>
<td>2.6.1</td>
<td>COMMUNICATION REQUIREMENTS</td>
<td>15</td>
</tr>
<tr>
<td>2.6.2</td>
<td>LOW INFRASTRUCTURE DEVELOPMENT AND MAINTENANCE COSTS</td>
<td>15</td>
</tr>
<tr>
<td>2.6.3</td>
<td>SCALABILITY</td>
<td>16</td>
</tr>
<tr>
<td>2.6.4</td>
<td>PRIVACY, INFORMATION SECURITY, AND NETWORK AVAILABILITY</td>
<td>16</td>
</tr>
<tr>
<td>2.7</td>
<td>SELECTION OF APPROPRIATE WIRELESS TECHNOLOGY AND ARCHITECTURE</td>
<td>16</td>
</tr>
<tr>
<td>2.8</td>
<td>EVALUATION OF PROPOSED COMMUNICATION ARCHITECTURE</td>
<td>17</td>
</tr>
<tr>
<td>2.8.1</td>
<td>EXPERIMENTAL SETUP</td>
<td>18</td>
</tr>
<tr>
<td>2.8.2</td>
<td>SELECTION OF ROUTING PROTOCOL</td>
<td>19</td>
</tr>
<tr>
<td>2.8.3</td>
<td>FEASIBLE OPERATING CONDITIONS</td>
<td>19</td>
</tr>
<tr>
<td>2.9</td>
<td>CONCLUSION</td>
<td>21</td>
</tr>
<tr>
<td>3.</td>
<td>CAPACITY ANALYSIS OF FEEDER-LEVEL LINEAR CHAIN NETWORK TOPOLOGY</td>
<td>22</td>
</tr>
<tr>
<td>3.1</td>
<td>INTRODUCTION</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>RELATED WORK</td>
<td>23</td>
</tr>
<tr>
<td>3.3</td>
<td>PROBLEM DEFINITION AND ASSUMPTIONS</td>
<td>24</td>
</tr>
<tr>
<td>3.4</td>
<td>SINGLE-CHAIN ANALYSIS</td>
<td>26</td>
</tr>
<tr>
<td>3.4.1</td>
<td>VALIDATION OF ANALYSIS THROUGH SIMULATIONS</td>
<td>28</td>
</tr>
<tr>
<td>3.4.2</td>
<td>PRACTICAL CONSIDERATIONS</td>
<td>29</td>
</tr>
<tr>
<td>3.5</td>
<td>MULTIPLE-CHAIN ANALYSIS</td>
<td>30</td>
</tr>
<tr>
<td>3.6</td>
<td>DATA RATE REQUIREMENTS ON COMMUNICATION INFRASTRUCTURE</td>
<td>31</td>
</tr>
<tr>
<td>3.6.1</td>
<td>AMI DATA OUTPUT</td>
<td>32</td>
</tr>
<tr>
<td>3.6.2</td>
<td>DATA TRANSFER INTERVAL FOR A GIVEN INFRASTRUCTURE</td>
<td>33</td>
</tr>
<tr>
<td>3.7</td>
<td>DATA HANDLING REQUIREMENTS FOR LARGE, HYBRID TOPOLOGIES</td>
<td>34</td>
</tr>
<tr>
<td>3.8</td>
<td>CONCLUSION</td>
<td>36</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS (continued)

Chapter | Page
--- | ---
4. SCALABLE COLLECTION OF METERING DATA IN SMART GRIDS THROUGH MESSAGE CONCATENATION | 37

4.1 Introduction | 37
4.2 Related Work | 40
4.3 Problem Statement | 42
4.3.1 Motivation | 42
4.3.2 Smart Metering Message Concatenation Problem | 43
4.4 Algorithms for SMMC Problem | 45
4.4.1 Nature of SMMC Problem | 45
4.4.2 SMMC Hardness Result | 46
4.4.3 Heuristic Algorithms | 48
4.5 Reference Algorithms | 52
4.5.1 EDF-Based Integer Linear Programming (ILP) Formulation | 52
4.5.2 Theoretical Optimal | 53
4.6 Evaluation | 54
4.6.1 Methodology | 54
4.6.2 Simulation Environment | 54
4.6.3 Distribution of Message Types | 55
4.7 Simulation Results | 55
4.7.1 Effect of Proposed Heuristic Algorithms | 56
4.7.2 Impact of Message-Type Distribution | 57
4.7.3 Impact of λ | 57
4.8 Impact of Network and Processing Delays | 59
4.8.1 Estimation of Network and Processing Delays | 59
4.8.2 Evaluation Results | 60
4.9 Data Volume with Lossy Links | 62
4.9.1 Theory | 63
4.9.2 Numerical Evaluation | 63
4.10 Conclusion | 65

5. SQUEEZE: AN EFFICIENT MULTI-LEVEL METER DATA-COLLECTION ALGORITHM FOR SMART GRIDS | 66

5.1 Introduction | 66
5.2 Background and Related Work | 69
5.3 Design Objectives and Problem Statement | 71
5.3.1 Motivation for Message Concatenation | 71
5.3.2 Network Topology | 72
5.4 Algorithms for MLDA Problem | 74
TABLE OF CONTENTS (continued)

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4.1 Heuristic Approach</td>
<td>74</td>
</tr>
<tr>
<td>5.4.2 Reference Algorithms</td>
<td>76</td>
</tr>
<tr>
<td>5.5 Evaluation</td>
<td>76</td>
</tr>
<tr>
<td>5.5.1 Theory</td>
<td>76</td>
</tr>
<tr>
<td>5.5.2 Simulation Environment</td>
<td>77</td>
</tr>
<tr>
<td>5.5.3 Message Types Distribution</td>
<td>77</td>
</tr>
<tr>
<td>5.6 Simulation Results</td>
<td>78</td>
</tr>
<tr>
<td>5.6.1 Effect of Proposed Heuristic Algorithms</td>
<td>79</td>
</tr>
<tr>
<td>5.6.2 Impact of Message-Type Distribution</td>
<td>79</td>
</tr>
<tr>
<td>5.6.3 Impact of Cost Functions</td>
<td>81</td>
</tr>
<tr>
<td>5.7 Conclusion</td>
<td>81</td>
</tr>
<tr>
<td>6. DISCUSSION, CONCLUSIONS, AND FUTURE WORK</td>
<td>83</td>
</tr>
<tr>
<td>6.1 Discussion</td>
<td>83</td>
</tr>
<tr>
<td>6.2 Conclusions</td>
<td>84</td>
</tr>
<tr>
<td>6.3 Future Work</td>
<td>85</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>86</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>95</td>
</tr>
<tr>
<td>A. Dissertation-Related Publications</td>
<td>96</td>
</tr>
<tr>
<td>B. Other Publications</td>
<td>97</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Comparison of end-to-end delay for varying node densities with data rates of 0.01 and 0.5 Mbps communication</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>Simulation results for different data rates</td>
<td>21</td>
</tr>
<tr>
<td>3.1</td>
<td>Communications for AMI at distribution level. The communications infrastructure needed (indicated by links shown as</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>clouds) can be divided into the following: (i) backhaul infrastructure communications possibly through feeders, and (ii) consumer/customer premise home area networks</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>Backhaul linear chain architecture considered in this work for AMI</td>
<td>25</td>
</tr>
<tr>
<td>3.3</td>
<td>Comparison of theoretical capacity and end-to-end throughput through simulations</td>
<td>29</td>
</tr>
<tr>
<td>3.4</td>
<td>Practical experiment results of ad hoc probe in on wireless multihop testbed</td>
<td>30</td>
</tr>
<tr>
<td>3.5</td>
<td>Impact of interfering flows on end-to-end chain throughput</td>
<td>31</td>
</tr>
<tr>
<td>3.6</td>
<td>Considerations for the selection of ideal sending interval for smart meters based on (a) housing density ( \rho_H )</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>and (b) chain capacity ( X_c )</td>
<td></td>
</tr>
<tr>
<td>3.7</td>
<td>Tree architecture of communication hierarchy</td>
<td>35</td>
</tr>
<tr>
<td>4.1</td>
<td>Data concentrator unit’s envisioned role of message concatenation at power distribution level</td>
<td>39</td>
</tr>
<tr>
<td>4.2</td>
<td>Smart meter datagram structure</td>
<td>42</td>
</tr>
<tr>
<td>4.3</td>
<td>Data concentrator schematic concept</td>
<td>46</td>
</tr>
<tr>
<td>4.4</td>
<td>Overall data reduction percentage using proposed heuristics over different message arrival rate and message type</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>distributions</td>
<td></td>
</tr>
<tr>
<td>4.5</td>
<td>Data reduction trend vs. delay addition</td>
<td>61</td>
</tr>
<tr>
<td>4.6</td>
<td>Average buffer size vs. delay addition</td>
<td>62</td>
</tr>
<tr>
<td>4.7</td>
<td>Data reduction savings vs. different backhaul technologies</td>
<td>64</td>
</tr>
<tr>
<td>5.1</td>
<td>Message concatenation at data concentrator units to reduce protocol overhead</td>
<td>67</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.2</td>
<td>Multi-level data concentration scenario</td>
<td>68</td>
</tr>
<tr>
<td>5.3</td>
<td>MLDA logical DCU 2-ary tree</td>
<td>72</td>
</tr>
<tr>
<td>5.4</td>
<td>Overall cost reduction percentage using proposed MLDA heuristics over different cost order functions and message type distributions</td>
<td>80</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Comparison of Possible Wireless Technologies</td>
<td>18</td>
</tr>
<tr>
<td>4.1</td>
<td>Types of Smart Meter Data Messages</td>
<td>43</td>
</tr>
<tr>
<td>4.2</td>
<td>Proposed Concatenation Heuristics</td>
<td>49</td>
</tr>
<tr>
<td>4.3</td>
<td>Pre-Defined Message Arrival Distributions</td>
<td>56</td>
</tr>
<tr>
<td>4.4</td>
<td>Heuristics Processing-Time Calculations</td>
<td>60</td>
</tr>
<tr>
<td>5.2</td>
<td>Proposed Multi-Level Aggregation Heuristics</td>
<td>75</td>
</tr>
</tbody>
</table>
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACORD</td>
<td>Appliance Coordination</td>
</tr>
<tr>
<td>AMI</td>
<td>Advanced Metering Infrastructure</td>
</tr>
<tr>
<td>ANSI</td>
<td>American National Standards Institute</td>
</tr>
<tr>
<td>AODV</td>
<td>Ad Hoc On-Demand Distance Vector</td>
</tr>
<tr>
<td>BBR</td>
<td>Beacon-Based Routing</td>
</tr>
<tr>
<td>BER</td>
<td>Bit Error Rate</td>
</tr>
<tr>
<td>BLR</td>
<td>Beacon-Less Routing</td>
</tr>
<tr>
<td>BPL</td>
<td>Broadband over Power Line</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>CBR</td>
<td>Continual Bit Rate</td>
</tr>
<tr>
<td>CFO</td>
<td>Cost Function Order</td>
</tr>
<tr>
<td>DAG</td>
<td>Directed Acyclic Graph</td>
</tr>
<tr>
<td>DCU</td>
<td>Data Concentrator Unit</td>
</tr>
<tr>
<td>DCU-OMA</td>
<td>Data Concentrator Unit-Optimized Message Allocation</td>
</tr>
<tr>
<td>DMS</td>
<td>Distribution Management System</td>
</tr>
<tr>
<td>DNP</td>
<td>Distributed Network Protocol</td>
</tr>
<tr>
<td>DSL</td>
<td>Digital Subscriber Line</td>
</tr>
<tr>
<td>EDF</td>
<td>Earliest Deadline First</td>
</tr>
<tr>
<td>EMU</td>
<td>Energy Management Unit</td>
</tr>
<tr>
<td>FEC</td>
<td>Forward Error Correction</td>
</tr>
<tr>
<td>FTTX</td>
<td>Fiber to the X</td>
</tr>
<tr>
<td>HAN</td>
<td>Home Area Network</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>IGF</td>
<td>Implicit Geographic Forwarding</td>
</tr>
<tr>
<td>ILP</td>
<td>Integer Linear Program</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MAC/PHY</td>
<td>Medium Access Control/Physical</td>
</tr>
<tr>
<td>MILP</td>
<td>Mixed-Integer Linear Program</td>
</tr>
<tr>
<td>MLDA</td>
<td>Multi-Level Data Aggregation</td>
</tr>
<tr>
<td>MSMMC</td>
<td>Multi-Level Smart Metering Message Concatenation</td>
</tr>
<tr>
<td>MTU</td>
<td>Maximum Transmission Unit</td>
</tr>
<tr>
<td>MV</td>
<td>Medium Voltage</td>
</tr>
<tr>
<td>NP</td>
<td>Non-Deterministic Polynomial-Time</td>
</tr>
<tr>
<td>NS2</td>
<td>Network Simulator Version 2</td>
</tr>
<tr>
<td>OMA</td>
<td>Optimized Message Allocation</td>
</tr>
<tr>
<td>PDF</td>
<td>Packet Delivery Fraction</td>
</tr>
<tr>
<td>PER</td>
<td>Packet Error Rate</td>
</tr>
<tr>
<td>PLC</td>
<td>Power Line Communication</td>
</tr>
<tr>
<td>PMU</td>
<td>Phasor Measurement Unit</td>
</tr>
<tr>
<td>PRIME</td>
<td>PoweRline Intelligent Metering Evolution</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>SMMC</td>
<td>Smart Metering Message Concatenation</td>
</tr>
<tr>
<td>SPR</td>
<td>Shortest Path Routing</td>
</tr>
<tr>
<td>STFG</td>
<td>Super Task Flow Graph</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>TCP/IP</td>
<td>Transmission Control Protocol/Internet Protocol</td>
</tr>
<tr>
<td>TPS</td>
<td>Total Packet Size</td>
</tr>
<tr>
<td>WAMPAC</td>
<td>Wide Area Monitoring, Protection, and Control</td>
</tr>
<tr>
<td>W-BPL</td>
<td>Wireless-Broadband over Power Lines</td>
</tr>
<tr>
<td>WiFi</td>
<td>Wireless Fidelity</td>
</tr>
<tr>
<td>WiMax</td>
<td>Worldwide Interoperability for Microwave Access</td>
</tr>
<tr>
<td>WSN</td>
<td>Wireless Sensor Network</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

According to the U.S. Department of Energy, “smart grid” generally refers to a class of technology people are using to bring utility electricity delivery systems into the 21st century, using computer-based remote control and automation. These systems are made possible by two-way communication technology and computer processing that has been used for decades in other industries. They are beginning to be used in electricity networks, from power plants and wind farms to the consumers of electricity in homes and businesses. They offer many benefits to utilities and consumers—mostly seen in large improvements in energy efficiency on the electricity grid and in the energy user’s homes and offices. For a century, utility companies have had to send workers out to gather much of the data needed to provide electricity. Workers read meters, look for broken equipment, and measure voltage, for example. Most devices that utilities use to deliver electricity have yet to be automated and computerized. However now many options and products are being made available to the electricity industry in order to modernize it.

The “grid” is comprised of networks that carry electricity from generating plants to consumers. It includes wires, substations, transformers, switches, and much more. The concept of smart grids can be thought of as “computerizing” the electric utility grid; however the difference is that communication capabilities come with the additional computation capabilities of devices associated with the grid. Each device on the network can be given sensors to gather data (power meters, voltage sensors, fault detectors, etc.) plus two-way digital communication between the device in the field and the utility’s network operations center. A key feature of the smart grid is automation technology, which allows the utility to adjust and control each individual device or millions of devices from a central location.

Electric power transmission systems typically operate above 110 kV, whereas electricity distribution systems operate at lower voltages. The distribution system automation requires adding communication capabilities over wires, switches, and transformers that connect the utility substa-
tion and control center to (smart) meters installed at the customer’s end. The need for improved communications at the power distribution level has taken on greater importance with the introduction of the smart grid approach. Title XIII of the Energy Independent and Security Act 2007 [1] requires improved operation of distribution systems. This includes development and incorporation of demand-side and energy-efficiency resources; deployment of real-time, automated, interactive technologies that optimize the physical operation of appliances and consumer devices; and provisions for timely information and control options to consumers, to name a few. These developments and deployments require additional capabilities of the grid, especially a better communication infrastructure beyond the supervisory control and data acquisition (SCADA) system.

For example, advanced metering infrastructure (AMI) initiatives are a popular tool to incorporate changes for modernizing the electricity grid, reduce peak loads, and meet energy-efficiency targets at the distribution system. With the introduction of AMI technology, two-way communication between a smart meter (SM) and the control center, as well as between the smart meter and customer loads would be facilitated for demand response, dynamic pricing, system monitoring, cold load pick-up, and greenhouse gas-emission mitigation [87]. AMI uses technology to capture and transmit energy use to a concentration point on an hourly or sub-hourly basis in contrast to standard meters that provide a daily energy usage total and a cumulative monthly bill. This application requires bidirectional communication: control commands from the control center of the utility to smart meters, and load profiles and logs from smart meters to the control center.

The critical requirement for advancement in the distribution system is thus real time information sharing and automation. By improving the communication infrastructure, a vital ingredient for the smart grid, a more reliable approach could be taken to better manage assets. In addition to asset and outage management tasks, communication will also aid in better energy management and tariff-related information. Deployment of smart distribution systems necessitates proper identification of power system requirements and integrating suitable communication and control infrastructure. The information communication and control layer of the smart grid brings about numerous advances, including the empowerment of customers to actively participate in the main-
tenance of a supply-demand balance around the clock and reliability improvements in electricity service.

Significant work has been done on power system communication needs and applications. IEC 61850 and DNP 3 standardize communication within the substation. ANSI C12.22 networking standards were built for advanced metering infrastructure [80]. Even though 80% of consumer interruptions are attributed to distribution component failure (at the feeder level), obtaining reliable information is still a challenging task. This is due to lack of component monitoring in the distribution system and the communication infrastructure. As a result, failure/abnormality analysis is done by harvesting information from the components at the substation level. A significant amount of work in analyzing such data [66, 17, 52, 8] has been done, but even investigating the entire feeder using the data from substations (beginning or end of a feeder) will not capture all the necessary information. Due to the non-trivial nature of failures/abnormalities, the exact prediction and location of a failure, independent of the feeder model, is still in the premature stages. By improving the communication infrastructure, a vital ingredient for the smart grid, a more reliable approach could be taken to better manage assets. In addition to asset and outage management tasks, communication will also aid in better energy management and tariff-related information.

With communications capabilities being an afterthought for the distribution system, it is not clear what communications technologies and architectures would help balance performance, cost, and efficiency. It is also not clear what would be the data-carrying capacity of specific technologies and architecture and how data flowing from different components can be carried to utility control centers through capacity or cost-constrained communication networks. Solving such challenges in communications network design are novel due to the following unique characteristics/constraints of smart grids:

- **Specific Purpose Applications, Data Characteristics, and Topology:** The nature of applications in the power grid results in a diversity of characteristics that sometimes need to be handled simultaneously by the communications network. For example, smart metering messages can be of different sizes with requirements to be delivered to their eventual destina-
tions in real-time, or tolerating some delay, or with no latency constraints. Communication network topologies resulting from the underlying physical topology of the power grid again dictates the use of some communications architectures/technologies that is typical in networks such as the Internet.

- **Privacy, Information Security, and Network Availability:** With applications like AMI fostering customer participation in energy delivery, ensuring privacy and information security in general is a challenge. With wireless technologies relying on a shared broadcast medium, unlike their wireline counterparts, this issue takes on greater dimensions. Further ensuring that the communication network is available for operation and control can be critical.

- **Scalability:** The scale of deployments possible under a single grid operator at the distribution level could be on the order of hundreds or thousands of nodes (even more if smart meters and end-user loads they connect to are included). Any wireless technology and the services it provides should be scalable in terms of the technical features it provides; such as data carrying capacity and latency.

- **Cost:** Bringing smart grid technology to the conventional grid requires many upgrades. Any communications technology/architecture used should be inexpensive, possibly unlicensed (to minimize spectrum licensing costs), be able to utilize existing infrastructure (like feeder poles), and be easy to set up and reconfigure. Preferably, the technology used should have low maintenance costs, including fees to third-party providers/carriers.

### 1.1 Contributions and Dissertation Organization

This dissertation is organized into three major parts corresponding to three primary contributions as follows:

**Communication Architecture for Power Distribution Systems:** The first part of this dissertation proposes a wireless mesh architecture for meeting the communication requirements be-
tween the control center and smart meters deployed in residences and commercial endpoints. This part considers wireless communication as a medium for feeder-level communication. It identifies the requirements for distribution feeder communication and explores the feasibility of wireless communication. Specific contributions of this first part are as follows: (i) encourage the design choice of using a wireless medium for communication at the power distribution feeder level, (ii) compare and contrast various wireless technologies, and identify a feasible subset for the envisioned architecture, (iii) evaluate how existing communication protocols designed for wireless mesh networks perform for network topologies expected in the power grid, and (iv) provide direction on how the wireless mesh architecture should be deployed to meet application requirements and optimize communication performance.

**Capacity Analysis for a Communication Architecture Used:** The second part of the dissertation studies the communication capacity of a specific communications architecture proposed for the AMI application. The contributions of this second part are as follows: (i) explore the feasibility of utilizing a wireless linear chain network to meet communication requirements for AMI, (ii) analyze the limitations imposed by such an architecture on the amount of data traffic that can be carried, and (iii) explore how grid operators could plan and operate their deployments.

**Data Volume Management:** In the third part, this dissertation tries to solve the data-gathering problem for a specific application like AMI given a communications architecture and network capacity constraints. This part introduces data concentrators or aggregators and their important role in reducing network capacity requirements through packet scheduling and aggregation policies applied along the data collection tree. Such algorithms and policies, however, do not exist currently and need to be developed, keeping in mind the unique characteristics of metering data like variable packet sizes, stochastic arrivals, and a combination of messages with deadlines and no deadlines. This part designs and comparatively evaluates a suite of online message concatenation algorithms at DCUs in the AMI scenario that
minimize usage of network capacity in transporting data through the meter data collection network while meeting quality-of-service (QoS) constraints imposed by applications on individual messages. The specific expected contributions of this part include the following: (i) a formulation of the message concatenation problem at data concentrator units (DCUs) in smart metering networks to minimize network capacity utilization, (ii) multiple heuristic online message concatenation algorithms that can be employed at DCUs for the message concatenation problem, and (iii) a comparative performance evaluation of proposed heuristic message concatenation algorithms.

This first part of this dissertation is presented as Chapter 2 and the second part as Chapter 3. The third part is split between chapters 4 and 5. Finally, Chapter 6 presents discussion of the contributions of this work along with general conclusions and possible future work.
CHAPTER 2
COMMUNICATION OPTIONS AT THE DISTRIBUTION LEVEL

This chapter provides the motivation for choosing to use wireless communication technologies over other options, and discusses possible limitations and the types of wireless architectures possible. This chapter sets the stage for further study into the communication performance of such architectures in the later chapters of this dissertation.

2.1 Introduction

The need for improved communication at the power distribution level takes on greater importance with the introduction of the smart grid approach. Significant work has been done on power system communication needs and applications. IEC 61850 and DNP 3 standardize communication within the substation. ANSI C12.22 networking standards were built for advanced metering infrastructure [80]. Even though 80% of consumer interruptions are attributed to distribution component failure (at the feeder level), obtaining reliable information is still a challenging task. This is due to lack of component monitoring in the distribution system and the communication infrastructure. As a result, failure/abnormality analysis is done by harvesting information from components at the substation level. A significant amount of work in analyzing such data ([8],[17],[52] and [66]) has been done, but even investigating the entire feeder using the data from substations (beginning or end of a feeder) will not capture all the necessary information.

Due to the non-trivial nature of failures/abnormalities, the exact prediction and location of a failure, independent of the feeder model, is still in the premature stages. By improving the communication infrastructure, a vital ingredient for the smart grid, a more reliable approach could be taken to better manage assets. In addition to asset- and outage-management tasks, communication will also aid in better energy management and tariff-related information.

The motivation for this dissertation was to propose a wireless mesh architecture for meeting the communication requirements between the control center and smart meters deployed in
residences and commercial endpoints. This work considers wireless communication as a medium for feeder-level communication. It identifies the requirements for distribution feeder communication and explores the feasibility of wireless communication. Specific contributions of this work include the following:

1. Encourages the design choice of using a wireless medium for communication at the feeder level.

2. Compares and contrasts various wireless technologies, and identifies a feasible subset for the envisioned architecture.

3. Evaluates how existing communication protocols designed for wireless mesh networks perform for network topologies expected in the power grid.

4. Provides direction on how the wireless mesh architecture should be deployed to meet application requirements and optimize communication performance.

2.2 Background and Related Work

Recently, many attempts have been made to deploy wireless technologies in a smart grid. Some of them have involved metering options and how to read their data, some have focused on sensor networks and receiving their data, and others have been based on feeder-level communication. The work of Laverty et al.[55] considers most last-mile options for telecommunication in a smart grid and discusses backhaul solutions for the distribution network. Recall from Section 2.1 that one approach for interconnecting a smart grid is using power line communication (PLC). Biagi and Lampe [13] considered PLC in low- and medium-voltage distribution grids to connect network nodes (e.g., meters, actuators, sensors) through multi-hop transmission. They investigated the application of geographic routing protocols and gauged their performance with respect to energy consumption and transmission delay. They investigated the use of beacon less routing (BLR), implicit geographic forwarding (IGF), and beacon-based routing (BBR). They also included shortest path routing (SPR) and flooding as benchmark schemes. In fact, they used greedy perimeter
stateless routing (GPSR) as a general geographic algorithm and BLR, IGF, and BBR as approaches of this algorithm to see which one achieved a performance close to that of SPR. What is remarkable in this paper is that SPR assumed perfect knowledge of instantaneous link qualities and relied on a centralized optimization. And BBR performed better than IGF, assuming that the frequency of hello messages was set commensurate with the network (connectivity) dynamics.

The connectivity of smart meters and their connectivity in smart grids is another subject that has been significantly researched. Zhao et al. in [90] proposed a unified solution for AMI integration with a distribution management system (DMS). They found that a challenge of the integration of AMI and DMS is that it entails different communication protocols and requirements for handling various meter information models. They claimed that by caching and delivering meter data back and forth between DMS and AMI systems, the proposed solution architecturally isolates the two systems, minimizes the influence of the AMI meter data load on DMS systems, and vice versa. Leon et al. proposed a two-layer wireless sensor network (WSN) for transmission towers, mainly to reduce the cost of operation while overcoming the limitations of a wireless communication range [56]. Muthukumar et al. proposed a WSN for distribution-level automation [68].

While the prior work above discusses how to design a network, some researchers have looked beyond. For example, Gormus et al. [36] discussed one of the key components of a future smart grid called load leveling, i.e., shifting the demand in time in order to match the available supply and in so doing improve the utilization of resources and reduces the reliance on environment-unfriendly reserve sources of energy as much as possible. The challenge here is in achieving such load leveling, and this work elaborates on how existing techniques from networking research could be potentially applied to solve these problems. Regarding connectivity of home-area networks to smart meters, Erol-Kantarci and Moufta [31] studied the connection of a home-area sensor network to an energy management unit. They proposed the appliance coordination (ACORD) scheme, which uses an in-home wireless sensor network (WSN) and reduces the cost of energy consumption. The cost of energy increases at peak hours; hence, reducing peak demand is a major concern.
for utility companies. With this scheme, they aimed to shift consumer demands to off-peak hours. Appliances use the readily available in-home WSN to deliver consumer requests to the energy management unit (EMU). EMU schedules consumer requests with the goal of reducing the energy bill. Souryal et al.[82] described an approach to modeling wireless communications at the link layer of the power grid. First, it identifies the various applications utilizing a specific link. Second, it translates the requirements of these applications to link traffic characteristics in the form of a link-layer arrival rate and average message size. Third, it uses a coverage analysis to determine the maximum range of the technology under an outage constraint and for a given set of channel propagation parameters.

Finally, using the link traffic characteristics and coverage area determined above, it employs a medium access control/physical (MAC/PHY) model to measure link performance in terms of reliability, delay, and throughput.

Some researchers believe that wireless communication is not enough to meet the entire needs of smart grid communication. Tsiropoulos et al.[86] describe a hybrid wireless-broadband over power lines (W-BPL) technology. They believe that this combination is suitable for rural and remote areas. The hybrid approach employs broadband over power lines (BPL) technology for the transmission of communication signals via the medium voltage (MV) grid and wireless technology for providing broadband access to end users. The authors showed the advantages and opportunities of this approach in a case study of Larissa, a rural area in central Greece. This network offers broadband access and smart grid applications along a 70 km MV power grid. In this dissertation, the focus is on feeder-level communication requirements and challenges. Having chosen the wireless communication medium for the grid, the goal here is to compare different technologies and pick one, considering the characteristics of the grid. After identifying a suitable technology, the best architecture for it will be determined and optimized for parameters like transmit power and receiver sensitivity of individual nodes, distance between nodes, protocol data rate, and other factors that specify the networks performance. The network performance is studied by first selecting
a suitable routing protocol and then use it as the basis for evaluating other parameters mentioned above.

2.3 Choice of Communication Medium

A communication network for the power grid allows power utility companies to access electricity usage data and services remotely, regardless of their geographic position. Real-time monitoring of transmission and distribution lines for protection against natural disasters or even malicious attacks are all reasons to have a secure, reliable, and scalable communication network for the power grid. Several last-mile options are available for getting a communication network to be operational; Broadband technologies like digital subscriber line (DSL), fiber Options such as fiber to the x (FTTX), and power line communication (PLC) are some examples. All these technologies have their limitations, however, due to their fixed nature and lack of flexibility.

When distribution feeders are considered, PLC is well-suited, because it is a no-cost medium for the utility and is spread along the distribution system. PLC has the potential to transmit data at a maximum rate of 11 Kbit/s, and the maximum data rate can be achieved only in a narrow frequency range of 9 to 95 kHz [10]. This low rate of communication is not enough for supporting applications where large amounts of data may be transferred, for example when large number of smart meters connected to end-user loads send periodic information using the AMI.

The current developments in the BPL could create an impression that this is the best technology, but that is not so. The distribution system is consistently affected by voltage transients and harmonics that are unpredictable, and is thus prone to a high level of disturbance. High-frequency signals involved in BPL need to bypass transformers to avoid high attenuation [85]. The attenuation in a radial distribution feeder is high, and this would increase the number of regenerators needed. It is expected that a typical 20-mile-long rural feeder needs regenerators on the order of 30 to 100 [24]. Thus, even if the medium of communication is free in BPL, there are infrastructure costs involved. In addition, the high-frequency signals may be blocked by voltage regulators, reclosers, and shunt capacitors, which are common in long radial feeders [24] posing problems.
Another option would be to use dedicated wired communication; however, one problem with copper wire connections is interference and attenuation. Fiber optic cables would be a solution to the interference, but would increase the cost. It should be noted that the investment required for a fiber optic network would be on the order of $10-100 million for 100 nodes [30]. Newly developing communities would be able to install a fiber optic communication network close to the feeders, enabling the infrastructure to be shared for both power grid and consumer communication needs. One of the advantages of this medium is that the utility would bear only the costs of the terminal equipment and for leasing the line, which would reduce the utility’s overhead and improve communication. On the other hand, the utility would not have control over the medium because, in most cases, it would not own the dedicated wired network and physical connections that reduce flexibility would be required.

All of the above technologies further have the disadvantage that when an electric pole falls down, it takes the communication link down as well. This would be a major concern when the communication is used for automatic fault location and system restoration, where communication is expected to help bring the power grid back to normalcy. For smart grid applications, a highly reliable communication network is necessary, with some prior work recommending an availability as high as 99.995% for the communication network [63]. This percentage requirement would result in the per-year unavailability of communication to less than 4.4 hours. All these concerns build the case to explore other options for the communication medium at the feeder level.

Wireless communication is a promising alternative for distribution-level communication. One of the important characteristics of wireless communication is the feasibility of communication without a physical connection between two nodes, thus ensuring continued connectivity even when a few poles are fallen down. In other words, redundant paths for communication are possible without additional cost. Another advantage of using wireless communication is that the utility must own only the terminal units, which are relatively cheap and could be integrated with cost-effective local processors. When multi-hopping is used in wireless communication, the range of communication could be extended, and the nodes located on the feeder would be able to communicate
with the control center. A major concern with a wireless medium is easy accessibility, which could result in security issues but could be avoided by using security mechanisms presented in prior work [16]. One disadvantage of wireless communication could be interference due to the presence of buildings and trees, which might result in multi-paths. Also, rural feeder sections could be long, and the range of communication might become a concern. These issues could be avoided to some extent with improved receivers and directional antennas. The issue of interference from power lines to wireless communication will be discussed next.

2.4 Impact of Interference due to Transmission Lines on Wireless Medium

One of the concerns in using wireless communication along power lines is the interference from high-voltage transmission lines. Electromagnetic noise generated around high-voltage power lines is an undesirable disturbance and can affect wireless data transmission. This noise can be observed as an additive signal to the original one, and it can interrupt, obstruct, degrade, or limit the performance of a communication system. According to Huertas and Echeverry [42], this noise is due to the following:

**Discharges between line components:** This occurs only in power lines less than 70 kV. This type of noise is generated in insulators, metallic parts, or faulty or improperly installed equipment. It tends to dominate the frequency spectrum between 10 and 20 MHz. Its effects can be controlled by ensuring a correct power line installation and proper maintenance.

**Corona effect:** This affects power lines over 110 kV and tends to dominate the frequency spectrum between 10 and 30 MHz. It is generated due to partial discharges in areas with a very high electric field and causes acoustic noise, radio interference, and mechanical vibrations.

Huertas and Echeverry [42] concluded that the radio interference generated by high-voltage lines diminishes logarithmically with the distance to the power line and with increasing frequency. Therefore, it is recommended that communication modules be operated at frequencies greater than 100 MHz. A selection of wireless communication technologies like Wi-Fi, ZigBee, or WiMax,
which operate in the GHz range, could be utilized in the distribution system with minimal interference.

2.5 Types of Wireless Communication Architectures

Wireless communication architectures could be classified based on whether they use single-hop or multi-hop communication between two endpoints, and whether they employ only a single technology or a combination of multiple technologies. Single-hop communication is possible only when the distance between the two points of interest is small enough to fit within the communication range of a technology. In many cases, when large geographical distances are involved, such as the case of the distribution system, multi-hop communication will be required where intermediate nodes forward data from the source to the destination. Wireless multi-hop communication provides several benefits over long-distance communication, like extending coverage due to multi-hop forwarding, greater throughput due to shorter hop distance, and possibly lower costs. A combination of wireless technologies could be employed in cases where one technology cannot cover a region, or does not have the capacity to support generated data traffic, or for economic reasons. Later in this dissertation, a multi-hop wireless communication architecture is proposed and evaluated by looking at the capacity limitations of using only a single technology, and how it scales up when using a combination of multiple technologies.

2.6 Wireless Communication Network Architecture for Distribution System

Having made a case for using wireless technologies in the previous section, some communication requirements for emerging applications at the distribution level are described here. The fault location application is one such example. Existing practices for locating possible faults at the distribution level involve many manual interventions. Only approximate locations of faults are known at the time of an outage, and operators need to spend time to identify the exact location, determine cause of the failure, and then fix it. If each feeder pole had a wireless master node, with the help of local sensors on power lines, it could provide useful information on the power
lines passing through it in a matter of seconds. Thus, the fault location and recovery could be much faster and lead to much more automation involving significantly less manual intervention. The AMI application would also benefit a wireless communication architecture making it easy to collect data from smart meters, requiring little infrastructure support like cabling, and providing flexible reconfigurations. To help realize such applications and make progress towards identifying communication performance, multiple candidate wireless technologies are compared, and the ability of linear chain architecture to meet some of these requirements are proposed and evaluated.

2.6.1 Communication Requirements

Using the two applications of AMI and automated fault location mentioned above as a guideline, and based on the general characteristics of the distribution system, the following is a list of requirements identified as important for any communication architecture to satisfy:

2.6.1.1 Low-Latency Communications

Any wireless technology should be able to provide low-latency communications from the data generation or collection point to the eventual destination. For an application like fault location, as soon as an abnormal state is sensed, this event should be communicated from the feeders to the control center for possible action. Any control commands from grid operators should similarly reach localized points on the grid with minimal delay. For applications like AMI, a higher latency is tolerable for the data collected from smart meters to the control center, but control commands in the other direction (from the control center to smart meters) for controlling loads and remote connects/disconnects need to be communicated immediately.

2.6.1.2 Low Infrastructure Development and Maintenance Costs

Modernizing the grid involves many upgrades. It is imperative that the cost of developing additional infrastructure is minimized by reusing existing infrastructure where possible. Thus, any wireless technology should be inexpensive, possibly unlicensed (to minimize spectrum licensing costs), able to utilize the existing infrastructure (like feeder poles), and easy to set up and re-
configure. The technology should also preferably have low maintenance costs, including fees to third-party providers/carriers.

2.6.1.3 Scalability

The scale of deployments possible under a single-grid operator at the distribution level could be on the order of hundreds or thousands of nodes (even more, if smart meters and end-user loads they connect to are included). Any wireless technology and the services it provides should be scalable in terms of the technical features it provides, such as data carrying capacity and latency.

2.6.1.4 Privacy, Information Security, and Network Availability

With applications like AMI fostering customer participation in energy delivery, ensuring privacy and information security in general is a challenge. With wireless technologies relying on a shared broadcast medium, unlike their wireline counterparts, this issue takes on greater dimensions. Further ensuring that the communication network is available for operation and control can be critical.

This dissertation mainly focuses on scalability of the communication backhaul and, to some extent, low-latency. The exploration of other requirements are left for future work.

2.7 Selection of Appropriate Wireless Technology and Architecture

The choices of wireless technologies—Wi-Fi, WiMAX, ZigBee and cellular data service—are compared in Table 2.1. Based on this information, and the fact that electric poles are typically separated by 100-300 feet\(^1\) (less than 100 meters), Wi-Fi seems to be the wireless technology that meets communication needs at the distribution level. Wi-Fi is based on unlicensed frequency bands and provides cost benefits and additional robustness by possibly leveraging existing community

---

\(^1\)This varies considerably based on terrain and population density. For example, in some terrain, the separation has to be large due to the inability to set up poles any closer together. In low-population density, rural areas, electric poles can also be found to have larger separations. In such cases, longer-range directional antennas might be needed, or long-range technologies like WiMAX must be used.
networks based on the same technology [6, 81, 88]. Among the unlicensed bands available for Wi-Fi, the 2.4-GHz range is most suitable due to its greater communication range when compared to higher-frequency bands. Though Wi-Fi has its security and privacy challenges, considerable progress has been made in recent years, including some specifically for applications like AMI that involve customers directly [16]. WiMAX is a good alternative technology that could be used as a gateway for long-haul communication based on availability and cost considerations. Data rates for ZigBee are too low to be a long-term solution in handling all the data expected at the backhaul as AMI penetration edges closer to 100%. The author of this work certainly believes that there is no one good solution that fits all scenarios and that utilities might deploy a combination of various options. In this dissertation work, the focus is on Wi-Fi and some of the apparent benefits it provides, as described above.

A Wi-Fi-based communication architecture for the scenario will behave in a multi-hop fashion where a source node relies on intermediate nodes on a routing path to forward packets toward the destination, with each data packet expected to pass through several hops. Each item (poles, smart meters, control center computers, etc.) can be a source, destination, or intermediate node to send, route, or receive a data packet. Also, each entity in these networks can be connected to more than one other entity. For example, a smart meter in a residence area can be connected to two or three distribution line poles in its vicinity, thus providing possibly multiple (linear) paths to a destination like the control center, in turn adding robustness. The scalability and ability to meet latency requirements of such linear chain wireless communication topologies will be tested next.

2.8 Evaluation of Proposed Communication Architecture

The feasibility of using a collection of cost-effective, fixed wireless nodes relying on Wi-Fi technology forming a linear chain network is evaluated here. This work bases the feasibility of communication architecture on the performance measurement of the end-to-end delay, and two other measurements of scalability — packet delivery fraction and node density.
Table 2.1: COMPARISON OF POSSIBLE WIRELESS TECHNOLOGIES

<table>
<thead>
<tr>
<th>Attribute</th>
<th>WiMAX</th>
<th>Wi-Fi</th>
<th>ZigBee</th>
<th>3G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>High</td>
<td>Medium</td>
<td>Low/Medium</td>
<td>High</td>
</tr>
<tr>
<td>Range (Single-Hop)</td>
<td>500-900 m 50km (LOS)</td>
<td>200-400 m</td>
<td>10-75 m, 1500 m for Zigbee pro</td>
<td>22 mi (35km)</td>
</tr>
<tr>
<td>Max. Data Rate</td>
<td>70 Mbps</td>
<td>54 Mbps</td>
<td>250 kbps</td>
<td>5 Mbps</td>
</tr>
<tr>
<td>Frequency Band</td>
<td>2-11, 10-66 GHz</td>
<td>2.4, 5 GHz</td>
<td>2.4 GHz, 915 MHz, 868 MHz</td>
<td>1.8 GHz, 2.5 GHz</td>
</tr>
<tr>
<td>Band License</td>
<td>Free and Licensed</td>
<td>Free</td>
<td>Free and Licensed</td>
<td>Licensed</td>
</tr>
<tr>
<td>Flexibility</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Robustness</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
</tr>
</tbody>
</table>

2.8.1 Experimental Setup

Simulations were carried out in the open-source network simulator ver. 2, ns2 [45], which allows abstraction of all communication protocols and their performance evaluation for different network topologies and configurations of various network traffic types. Nodes were placed as a linear chain topology (mimicking electric poles on a distribution line) on a 10-km-long scenario, each node separated from the next by a distance of 100 meters. The distance between nodes was varied only if the impact of node density was studied, in which case the density (nodes per unit distance) was explicitly mentioned. In the topology, the first node from the left was assumed to be the source, while the destination was the right endpoint of the chain to which information was being sent. Continuous bit rate data traffic was used from the source with varying data rates in our experiments. For all simulations, a fixed antenna model was used with the two-ray propagation channel model. The transmit power was set to be fixed at 0.28 Watt, providing a range of 100 meters in the simulator, which is consistent with practical values for Wi-Fi. The channel data rate was set to 11-Mbps. The following performance measures were used for the routing protocol:

- **Packet delivery fraction (PDF):** This is the ratio of packets delivered to packets sent by the traffic generator.
End-to-end packet delay: This is the average of the delays encountered by all successfully received packets at the destination from the source node.

Node density: This term is used to describe the constant number of nodes deployed within a fixed geographic length.

2.8.2 Selection of Routing Protocol

Ad hoc routing protocols used for the mesh architecture can be divided into two main categories: reactive (on-demand) and proactive (table-driven). Other possible categories include location-based routing [54] and prediction based routing [70]. Extensive prior work on evaluating routing protocols for ad hoc networks[15]. The ad hoc on-demand distance vector (AODV) routing protocol was chosen [18] as the representative routing protocol in this work due to prior evaluation results and comparisons with proactive routing protocols. In on-demand routing protocols, routes are created as and when required. When a source wants to send to a destination, it invokes route-discovery mechanisms to find the path to the destination. These discovered routes time-out after a fixed duration, requiring new routes to be created to replace them. For the evaluations, some default parameters of the AODV protocol are modified to work with route lengths expected to be many hundreds of hops. For example, the default network diameter for the AODV in ns2 is 30 hops; then it is modified to be larger than the number of hops expected in the long chain topology.

2.8.3 Feasible Operating Conditions

As a first step, it was decided to determine the appropriate node density when deploying the proposed wireless communication architecture. Data rates of 0.01 and 0.5 Mbps were used. The end-to-end network delay for varying node density is shown in Figure 2.1. As node density increases initially, multiple nodes act as intermediate nodes and interfere with each other. The drop seen for node densities of 20 to 21, shown in Figure 2.1, is due to a sudden decrease in the number of hops taken by the AODV routing protocol from the source to the destination. This protocol considers all possible paths from the source to the destination and chooses the one that
can reach the destination with the least delay, which typically is the route composed of the shortest number of hops. As node density increases, there is a certain point at which the routes taken can skip over some nodes on the path to the destination. From Figure 2.1, it can be seen that node density of 10 nodes per km, which is also the minimum needed for end-to-end connectivity for the transmit power level used, performs better for both 0.01 Mbps and 0.5 Mbps cases. The lesser the node density, the cheaper the cost of deploying the communication architecture, because it reduces the number of nodes that need to be deployed. Therefore, this work suggests using the minimum possible node density for a given communication range of a technology.

![Figure 2.1: Comparison of end-to-end delay for varying node densities with data rates of 0.01 and 0.5 Mbps communication.](image)

Further analysis was done for delay and packet delivery for data rates between 0.01 and 1 Mbps. To compare the effect of node densities, both 10 and 15 nodes per km were simulated. Figure 2.2 shows the simulation results for these cases. From Figure 2.2(a) and (b), it can be seen that when the data rate exceeds 0.15 Mbps, the delay increases significantly, and the packet delivery fraction decreases from 100%.

From the simulations, it can be seen that even with the ideal communication scenario (no interference and only one node sending information), there is a limit on the capacity. In subsequent sections, these limits of a wireless communication architecture will be explored, and insights on planning deployments with possibly a combination of technologies will be offered. Readers inter-
ested in additional simulation based evaluation for the Wi-Fi linear chain topology are referred to the work of [46].

Figure 2.2: Simulation results for different data rates.

2.9 Conclusion

Several technologies have been suggested to meet communication needs at the distribution level. This work compared these technologies and provided a justification of why wireless communication technologies may be most suitable. A linear chain wireless communication architecture was proposed and evaluated for its ability to meet communication requirements of scalability and latency.
CHAPTER 3
CAPACITY ANALYSIS OF FEEDER LEVEL LINEAR CHAIN NETWORK TOPOLOGY

In this part of the dissertation, the capacity of a linear chain topology likely to be used for feeder level communications at the distribution level is theoretically analyzed. Although a Wi-Fi architecture was proposed as the technology of choice for communications at the distribution level in previous work [46], the treatment in this section of the dissertation and following sections is more general, allowing for the use of other wireless technologies in any communication architecture that is used.

3.1 Introduction

There are many communication technology and architecture options in deploying the AMI. As shown in Figure 3.1, the end-to-end communication infrastructure can be divided into two major phases: (i) consumer-level home area network (HAN) between a smart meter and various electrical equipment/appliances, (ii) the backhaul link that collects information from smart meters and carries it to the utility control center. The backhaul can be either a direct link from individual smart meters to the utility control center, or a link (or series of links) from a concentrator node, which aggregates data from multiple smart meters through some form of a mesh network to the control center. There is reasonable consensus for using ZigBee-based star topologies in HANs [7, 12], and ZigBee- or Wi-Fi-based mesh topologies to collect and aggregate data at concentrators [53], but the technologies and topologies used for the backhaul are still an open problem.

This work thus looks also at the design of backhaul communication for the AMI. An approach of deploying the backhaul communication infrastructure is to create a network through the existing feeder infrastructure. This approach has the advantage of reducing costs by reusing existing infrastructure of electric poles and feeder lines if needed. Further, it can also integrate well with sensors deployed on feeder poles to improve distribution automation through applications.
like automated fault location and self-healing feeders. It is not yet clear though what would be the communication capacity of a very long linear chain network deployed for the AMI application between concentrator nodes and the utility control center. The AMI application presents the challenge of collection and management of data from smart meters. By current standards, each smart meter sends a few kilobytes of data every 15 minutes to a smart meter [27, 74]. When this is scaled up to large numbers, many existing communication architectures will find it difficult to handle the data traffic due to limited bandwidth. To this end, this work explores the feasibility of utilizing a wireless linear chain network to meet communication requirements for AMI, analyzes the limitations imposed by such an architecture on the amount of data traffic that can be carried, and explores how grid operators could plan and operate their deployments.

Figure 3.1: Communications for AMI at the distribution level. The communications infrastructure needed (indicated by links shown as clouds) can be divided into the following: (i) backhaul infrastructure communications possibly through feeders, and (ii) consumer/customer premise home area networks.

3.2 Related Work

The work in [64] is closest to our work by considering a wireless backhaul architecture. Their focus is however on enhancements to the air interface and network protocols instead of
network capacity. The work in [55] considers most last-mile options for telecommunication in a smart grid, especially backhaul solutions for the distribution network. The authors of [13] considered PLC in low- and medium-voltage distribution grids to connect network nodes (e.g., meters, actuators, sensors) through multi-hop transmission. The authors in [90] proposed a unified solution for Advanced Metering Infrastructure (AMI) integration with a Distribution Management System (DMS). They found that a challenge of the integration of AMI and DMS is that it entails different communication protocols and requirements for handling various meter information models.

While the prior work above discusses how to design the communication network at the distribution level for AMI, some researchers have looked beyond. For example, [36] discussed one of the key components of a future smart grid called load leveling and elaborates how existing techniques from computer networking research could be potentially applied to solve these problems. The authors of [82] described an approach to modeling wireless communications at the link layer of the power grid, with emphasis on employing a medium access control/physical layer model to measure link performance in terms of reliability, delay, and throughput. Some researchers believe that wireless communication is not enough to meet the entire needs of smart grid communication. For example, the authors in [86] describe a hybrid Wireless-Broadband over Power Lines (W-BPL) technology. They believe that this combination is suitable for rural and remote areas.

### 3.3 Problem Definition and Assumptions

Consider a linear chain communication network that is multi-hop in nature with \( n + 1 \) nodes (\( n \) hops from source to destination) overall separated by a constant inter-node distance. Let \( C \) bps be the maximum single hop throughput or capacity, and \( r \) meters be the communication range of the technology used. The source node generates data at a rate of \( \lambda \) bits per second. Let the sink node be \( L \) meters away from the data concentrator. We would like to determine what is the end-to-end data capacity, \( X_{ce} \), of this linear chain topology.

This problem is representative of the type of linear chain network that will be used at the distribution level of the power grid as described in our proposed Wi-Fi architecture in the
previous section. The source node could represent a concentrator or data aggregator where all data from many distributed sources (e.g. smart meters of residences in the AMI application) could be collected and sent over the network. The sink node could represent a control center of the grid operator where all data is collected. The scenario could also be the other way around, where the source node is the control center sending commands to a sink node that could be an end-user’s smart meter. The smart meter to control center direction is more interesting in terms of network capacity analysis as it is expected to have more data; the other direction is mainly envisioned for control commands. Figure 3.2 shows an illustration of the specific application scenario at the feeder-level. The assumption of constant inter-node distance is based on the assumption that nodes are co-located with existing utility infrastructure (e.g. distribution feeders) that is deployed at some fixed density. Determining the maximum data traffic rate that can be supported by such a linear chain network for any given communication technology that is deployed is useful in network capacity and operational planning, and reliability analysis.

![Figure 3.2: The backhaul linear chain architecture considered in this work for AMI.](image_url)

Prior research on the capacity of ad hoc wireless networks have studied similar problems, which also include single chain capacity analysis [58, 4]. However, their focus was solely on Wi-Fi technology and considered many other topologies (apart from linear chain), including random
networks; our goal is to analyze and study the capacity of linear chain wireless topologies and how it applies to the distribution level, and without being restricted to only one technology.

### 3.4 Single-Chain Analysis

The shared broadcast nature of the wireless medium typically\(^2\) dictates that when one node is transmitting, none of the nodes within the interfering range of the transmission can transmit or receive. Let us assume that only one out of every \(m\) consecutive links of a linear chain can be active at a time. Therefore, a rough estimate of a single chain capacity is \(1/m\) times the one-hop capacity. This estimate may, however, be an over-estimate, because the technology used may not have perfect scheduling (true for example in the case of Wi-Fi at the MAC layer), and not all opportunities to transmit may be taken; it could happen that if the first link is active, any of the links \(m + 1\) to \(2m\) could be active simultaneously, if they have data to forward (link \(m + 1\) being active is optimal in this case). This estimate can be improved through the following analysis.

Assume a linear chain with \(n\) links, \(L_1\) to \(L_n\) with \(n\) sufficiently large. Let \(f(n)\) denote the average number of links that are active in the chain, given that the first link \(L_1\) is active. Let \(g(n)\) denote the average number of active links in the chain. A use a recursive analysis is used to calculate \(f(n)\) and \(g(n)\).

If \(L_1\) is active, then none of the links \(L_2\) to \(L_m\) can be active. Furthermore exactly one of \(L_{m+1}\) to \(L_{2m}\) will be active because nodes on the chain always want to send data, and any link \(L_i\) can be interfered with only by links \(L_{i-(m-1)}\) to \(L_{i+(m-1)}\). If \(L_{m+i}\) is active, then the number of active links in the chain will be \(1 + f(n - (m + i) + 1)\). Ignore the corner effect of the end of the chain where there may not be as many links interfering; assuming \(n\) is large supports this decision. If \(p_i, 1 \leq i \leq n\) is the probability that any link \(L_i\) is active, then the average number of active links given that \(L_1\) is active can be computed as

\(^2\)If techniques like code division multiple access are used, multiple nodes within interference range of each other can transmit at the same time. These are typically not used except in cellular systems, one reason being the complexity of the hardware involved.
\[ f(n) = 1 + \sum_{j=1}^{m} (p_j \cdot f(n - (m + j) - 1)) \]

Next, compute \( g(n) \), letting \( q_i, 1 \leq i \leq n \) denote the probability that \( L_i \) is active. Then,

\[ g(n) = \sum_{i=1}^{m} (q_i \cdot f(n - i + 1)) \]

where it is assumed that any link \( i \) from the first \( m \) links can be active, thus removing the conditionality imposed by \( f(n) \) on \( L_1 \) being active. Continuing,

\[ g(n) = \sum_{i=1}^{m} \sum_{j=1}^{m} (p_j \cdot f(n - i + 1 - (m + j) + 1)) \]
\[ = \sum_{i=1}^{m} q_i \cdot \sum_{j=1}^{m} (1 + f(n - i + 1 - (m + j) + 1)) \]
\[ = \sum_{i=1}^{m} \sum_{j=1}^{m} q_i \cdot p_j + \sum_{j=1}^{m} p_j \sum_{i=1}^{m} (q_j \cdot f(n - i + 1 - (m + j) + 1)) \]
\[ = \sum_{j=1}^{m} p_j \cdot g(n - i + 1 - (m + j) + 1) \]

Assuming that \( g(n) \) is a linear function of \( n \) and using \( an + b \) as a possible solution with \( b = 0 \), and \( p_i = 1/m, 1 \leq i \leq m \), it is possible to solve and obtain \( a = \frac{2}{3m-1} \). Thus, \( g(n) = an = \frac{2n}{3m-1} \) is the average number of links active in the linear chain at a steady state when \( n \) is large. This results in an overall chain capacity, \( X_c \), of

\[ X_c = \frac{g(n)}{n} C = \frac{2C}{3m-1} \] (3.2)

where \( C \) is the single-hop capacity. For smaller values of \( n \), the recursion in equation 3.1 can be utilized to compute different values of \( g(n) \) and compute the end-to-end chain capacity.
3.4.1 Validation of Analysis through Simulations

To verify these assumptions, a series of simulations was conducted and compared the results of the theoretical result above compared with simulations of the Wi-Fi technology. Note that other wireless technologies could be compared similarly by introducing correct parameter values for those technologies. The simulations were done with the ns2 simulator [45] whose parameters were set to a 2 Mbps channel data rate. The average end-to-end chain throughput is shown along with the 90% confidence intervals. Additional details of the simulation setup were described in 2.8.1. For the theoretical results, the value of $C$ in equation (3.1) is set to 1.7 Mbps, because overhead reduces useful throughput from the 2 Mbps channel rate. The communication range $r$ was set constant to 100 m, and the interfering range of each node was set such that $m = 4$ in the above analysis. The distance $L$ was varied to alter the chain length. The corresponding values of $g(n)$ were computed using equation (3.1) and the resulting capacity values for each chain length plotted.

From Figure 3.3, the results of the theoretical analysis and simulations can be compared. For the largest data packet size of 1,500 bytes, which is expected to have the greatest throughput, it can be seen that the difference between the two results is very small ($< 6\%$), validating the analysis. It is possible to further improve on the analytical result above (and get closer to the simulation results) by adding a correction for the way the IEEE 802.11 standard medium access control (MAC) protocol works for Wi-Fi. Under this protocol, there could be instances where none of the $m$ consecutive links could be transmitting, because they might be counting down backoff slots in accordance with the binary exponential backoff algorithm used [2]. This is not taken into consideration in order to keep the analytical result presented in this section more general.

---

3 Although current technologies have higher data rates available, the actual value of the data rate used does not change the overall result. Setting a data rate of 2 Mbps further proves useful in order to compare against prior work done in the area of capacity of ad hoc networks [58], where a similar analysis was done for random networks.

4 Figure 3.3 also shows confidence intervals of 90%, but this is not visible in the plots due to small intervals ranging from 6.92682E-05 Mbps to 0.000293148 Mbps during 12 runs of simulation.
3.4.2 Practical Considerations

Another step to confirm the results would be to perform experiments on a large-scale testbed. However, this approach would have the limitation that it would not be possible to scale the network up to gather any meaningful results. For large networks (or long chains), simulations are the best way to gauge performance. The NS2 simulator has been used for such large-scale simulations many times and provides a useful check against the analytical results. Comparing both simulation and analytical results provides a good validation check and is common in prior research with large-scale networks. Having said that, related work has considered the capacity of multi-hop wireless networks [[26], [49]], but the most relevant practical scenario to the architecture in this dissertation is presented in the work of Sun et al.[83], where the authors implemented a testbed to validate the path capacity on multi-hop fixed-rate wireless networks. The testbed included several 802.11b laptops kept about 70 ~ 80 meters (which is similar to the distance between power distribution poles) apart in a chain topology. For comparison, their results are shown in Figure 3.4. Remarkably, their experimental results are very similar to the results in this work, adding confidence to the practicality of the results here.
Figure 3.4: Practical experiment results of ad hoc probe [83] on wireless multihop testbed.

### 3.5 Multiple-Chain Analysis

If the rate of data generation is higher than what a linear chain communication topology can support, then multiple-chains could possibly be deployed. However, it is important that these chains do not interfere with each other’s flows. It is useful at this point to study the impact of interfering chains (or interference in general) on the end-to-end throughput achieved. A varying number of interfering chains 10 hops length and each hop spanning 100 m were simulated as before. The source node of each chain was configured to send at a rate of 100 kbps. The average end-to-end chain throughput is shown along with 90% confidence intervals in Figure 3.5. As can be seen, each interfering flow results in a rapid decrease in end-to-end throughput. Thus, interfering flows need to be carefully planned. The interfering flow chain could be co-located using the same physical path and the same (or different) nodes, but possibly different non-interfering frequency channels, or a different non-interfering path along another set of feeders.

After ensuring that flows are non-interfering, it is useful to extend the single-chain analysis to how many chains would be needed for some given data generation rate. The eventual wireless communication architecture could thus be composed of multiple independent linear chains from many concentrators to the control center. The presence of multiple such chains, if planned and deployed properly, can add some reliability to the communication network by providing the option of alternate paths on reconfiguration (from a non-overlapping non-interfering configuration).
Let $X_c$ be the capacity of a single chain, as found in equation (3.2). Let $N$ be the ratio of data to be sent from the concentrator, $\lambda$, to the capacity of chain $X_c$. In order to estimate the number of chains needed to support the required $\lambda$, the following relation is always true:

$$\lambda \leq NX_c$$  \hspace{1cm} (3.3)  

Thus, to obtain the number of required chains, the function $f$ is defined as:

$$f\left(N = \frac{\lambda}{X_c}\right) = \begin{cases} 
1 & \text{if } N \leq 1 \\
\lceil N \rceil & \text{if } N > 1 
\end{cases}$$  \hspace{1cm} (3.4)  

### 3.6 Data Rate Requirements on Communication Infrastructure

The analysis in the previous section looked at how much data rate, $X_c$, a single linear chain communication network can support from a source to sink and how many chains, $N$, will be needed to carry a specified amount of data traffic, $\lambda$. In this section, an analysis of data communication requirements when using the AMI application will be performed in order to better understand the magnitude of $\lambda$ that must be supported. The application scenario under consideration was shown previously in Figure 3.2, where the smart meter aggregator or concentrator point gathers data from
smart meters and sends it to the control center. This section will also provide insights into the amount of data and interval at which smart meters should send this data for a given communication infrastructure.

### 3.6.1 AMI Data Output

Suppose there are \( z \) smart meters in a given area as part of the AMI application, all connected to a concentrator. Because the rate of data generation by smart meters is not necessarily synchronized, an average rate of data generation per second is a more useful value. Let each meter generate a data packet \( P \) bytes long at an interval of \( t \) minutes. Thus, the average data traffic rate reaching the concentrator from \( z \) smart meters will be

\[
\lambda = \frac{z \cdot P}{60 \cdot t} \text{ bytes/sec} = \frac{z \cdot P}{7.5 \cdot t} \text{ bits/sec} \quad (3.5)
\]

Current standards specify that each smart meter sends a 512-byte packet every 5, 15, 30, or 60 minutes[27], but 15 minutes is most common [74]. Thus, henceforth we use the interval of 15 minutes for evaluations. This translates to a data rate of \( \lambda = 4.55z \) bps arriving at the concentrator.

This simple analysis can be further extended to include the concept of smart meter density. For example, assume each house in a residential neighborhood has a smart meter, and the smart meter density (equivalent to housing density) per square meter, say \( \rho_H \), is known. Therefore, given the area of the location of interest, \( A \), and the housing density, \( \rho_H \), the expected data traffic arriving at the concentrator would be

\[
\lambda = \frac{\rho_H A \cdot P}{60 \cdot t} \text{ bytes/sec} = \frac{\rho_H A \cdot P}{7.5 \cdot t} \text{ bits/sec} \quad (3.6)
\]

Figure 3.6(a) plots the data traffic at the concentrator for various values of \( \rho_H \) for an area of 1,000 sq. m and packet size \( P = 512 \) bytes. Three intervals at which meters could send data

---

5Technologies used in gathering data from each individual customer to the smart meter and the capacity analysis of this network is not addressed in this work and would be part of future work.
are considered. It can be seen that huge amounts of data can be easily generated by applications in dense AMI deployments.

![Diagram showing data concentrator data rate and minimum sending interval for various time intervals](image)

(a) Incoming data rate $\lambda$ at concentrator for various $t$ and $\rho_H$

(b) Minimum smart meter sending interval $t$ allowed for a given communication infrastructure capacity $X_c$.

Figure 3.6: Considerations for the selection of ideal sending interval for smart meters based on (a) housing density $\rho_H$ and (b) chain capacity $X_c$

### 3.6.2 Data Transfer Interval for a Given Infrastructure

If the communication infrastructure is already in place, then the parameter that decides the data rate that will need to be supported by this infrastructure is the interval $t$ at which each smart meter sends data given $\rho_H$, $A$, $P$, $N$, and $X_c$. Equations (3.3) and (3.6) can be used to the following inequality:

$$\frac{\rho_H \cdot A \cdot P}{7.5 \cdot t} < N \cdot X_c$$

which simplifies to provide a lower bound for the sending interval as

$$t > \frac{\rho_H \cdot A \cdot P}{7.5 \cdot N \cdot X_c}$$

(3.7)
Figure 3.6(b), shown previously, provides an idea of how the sending interval for each smart meter increases with a decrease in end-to-end capacity of the communication infrastructure for $m = 4$, $N = 1$, $A = 1000$ sq. m, and $P = 512$ bytes.

### 3.7 Data Handling Requirements for Large, Hybrid Topologies

The analysis thus far has been based on a homogeneous architecture where only one type of technology was used to transport data from concentrators to the control center. Here, the analysis is extended to study the expected data traffic that needs to be handled over a large, wide-area network where possibly multiple technologies might be needed from concentrators to reach the control center. For example, one technology could be used within metropolitan areas due to a higher population density, while another technology is used for suburban or rural users. The analysis begins by developing a tree-based model for the communication infrastructure and subsequently using this model for assessing data traffic handling requirements.

Figure 3.7 shows how a tree architecture can be used to model $h$ different levels of technologies used end to end. There is a root node, which is the control center, and the final destination of all data collected from smart meters. All leaf nodes at the lowest level of the tree are smart meters generating data. All other intermediate nodes of the tree that are neither leaves nor the root node, are the concentrators. This work is interested in the data traffic at each concentrator and the overall data traffic received at the control center.

Let the height of the tree be $h$, starting from level $0$ to level $h$. Let there be $n_c(i)$ cluster(s) at each level $i$. Then there are $N(i, k)$ aggregators (concentrators) in a $k$th cluster of level $i$, where $0 < k < n_c(i)$ and $i > 0$. This helps to identify the lower-level data traffic senders to which a concentrator $j$ of cluster $k$ at level $i$ is connected. Thus, $\lambda_{ijk}$ is the rate of data collected at concentrator number $j$ in cluster $k$ at level $i$ and defined as

$$\lambda_{ijk} = \sum_{j' = 1}^{N(i', k')} \lambda_{ij'k'}$$  \hspace{1cm} (3.8)
where $i' = i - 1$ (lower level), $j'$ is the aggregator index in cluster $k'$, and $k'$ is the cluster identifier in level $i'$ that sends its data to aggregator with identifiers $i, j,$ and $k$.

In this dissertation, to represent the total data traffic that is being collected at any specific level $i$, $\lambda_{i..}$ is used and defined as

$$\lambda_{i..} = \sum_{j=1}^{n_{c(i-1)}} \lambda_{ijk}$$

(3.9)

where $\lambda_{111}$ is a base case and is the data traffic by an aggregator at level 1 directly from smart meters (leaves at level 0) connected to them and equals

$$\lambda_{111} = \sum_{j=1}^{N(0,1)} \frac{P_1}{7.5 \cdot t} \text{ bps}$$

(3.10)

based on the data generation rate for each smart meter from equation (3.5) (without the multiplier $z$) for all concentrators $j$ at level 1, and $P_k$ is the packet size used for data sent in cluster $k$.

Finally, the overall data traffic that is being transmitted from smart meters to the root control center over the distribution grid is equal to
\[ \lambda_{\text{Total}} = \sum_{j=1}^{N(0,k)} \sum_{k=1}^{n(0)} \lambda_{0jk} + \sum_{j=1}^{N(1,k)} \sum_{k=1}^{n(1)} \lambda_{1jk} + \cdots + \sum_{j=1}^{N(h,k)} \sum_{k=1}^{n(h)} \lambda_{hjk} \]

\[ = \sum_{i=0}^{h} \sum_{j=1}^{N(i,k)} \sum_{k=1}^{n(i)} \lambda_{ijk} \quad (3.11) \]

### 3.8 Conclusion

In this section of the dissertation, a theoretical study was done to understand the capacity limitations of using linear chain wireless technologies at the distribution level. The application of advanced metering infrastructure that encourages customer participation in energy delivery was used as a case study to understand the implications of any limitations imposed by the proposed communication architecture.

Although one specific communication architecture for the distribution level of the power grid was recommended in this dissertation, an optimal architecture may vary widely. For example, at the distribution level of the power grid, engineers could choose to utilize a configuration in which short-range wireless communication technology is used to send data from sensors installed on electric poles to a master node on the pole (so that no wire is needed between the sensors and router), and then wired technologies such as DSL, FTTX, or PLC are used to feed the data back to the control center. Alternative configurations may allow smart meters to send data directly back to the collection point without involving any wired technology. In response to such a wide range of possible network and system configurations, the best practice begins with a thorough and careful system architecture and configuration design. After this process, several distinct configurations that involve wired or wireless technologies of varying degrees may be derived. Therefore, apart from proposing a specific architecture based on some common applications, this dissertation also contributes to a step in the direction of providing a suite of communication topologies and an analysis of their technical feasibility, an optimal configuration, and deployment considerations.
CHAPTER 4
SCALABLE COLLECTION OF METERING DATA IN SMART GRIDS
THROUGH MESSAGE CONCATENATION

4.1 Introduction

Advanced metering infrastructure uses technology to capture and transmit energy use to a
collection point on an hourly or sub-hourly basis in contrast to standard meters that provide a daily
energy usage total and a cumulative monthly bill [28]. With the introduction of AMI technology,
two-way communication between a “smart” meter and the grid operator’s control center, as well
as between the smart meter and consumer appliances, would be facilitated for demand-response,
dynamic pricing, system monitoring, cold-load pick-up, and greenhouse gas-emission mitigation
[87]. The information communication and control layer of the smart grid brings about numerous
advances, including the empowerment of customers to actively participate in the maintenance of
the supply-demand balance around the clock and the resulting reliability improvement in electricity service. There are many benefits to grid operators, consumers, and society as a whole from adopting AMI technologies [72].

In addition to AMI, many other applications will be enabled by information flow across
the electric power grid, including distributed generation, state estimation of the power distribution
system, and demand-side management, to name a few. A big challenge for smart grid application
scenarios, and the information-sharing framework that enables them, will be handling the massive
amount of data that is expected to be collected from data generators and sent through the commun-
ication backhaul to the grid operator. For example, by current standards, each smart meter sends a
few kilobytes of data every 15-60 minutes to a smart meter [27, 74]. When this is scaled up to many
thousands, existing communication architectures will find it difficult to handle the data traffic due
to limited network capacities, especially in limited bandwidth last-mile networks [5, 62]. Future
applications may require data to be collected at a finer granularity, thus adding to the challenge
[29]. Network capacity is a precious resource for electric utilities because they are either leasing
such networks from third-party providers [43], or building infrastructure themselves and leasing bandwidth out (especially at the backhaul) to recuperate investment costs [51]. In either case, it is in the interest of electric utilities to reduce the volume of information transported through these networks for smart grid applications while ensuring that application QoS requirements are met.

One approach to reducing the volume of data given some application sampling rate, is to concatenate multiple messages into a larger packet to reduce protocol overhead due to packet headers. This approach has the potential to reduce network capacity requirements significantly (quantified later in this dissertation) due to the small size of messages sent in smart metering networks, with packet headers possibly being of a comparable size to the underlying message sent. Such concatenation of messages can be done by each smart meter itself. However, each meter may not generate messages frequently enough to be able to have the chance to concatenate enough packets to reduce overhead significantly and also meet their stated application deadlines. Each meter is also expected to be relatively constrained (compared to a concentrator) in terms of data storage capabilities to keep a large window of packets from which to aggregate. Thus, a better approach is to concatenate messages at an intermediate point upstream from individual meters.

Such an intermediate point where message concatenation can be done is at data concentrator units (or a similar entity, sometimes called a data aggregator) that collect data from many smart meters and forward them upstream. Figure 4.1 depicts this concept and shows the DCU’s role at the power-distribution level of the power grid. Data concentrators or aggregators can play an important role in reducing network capacity requirements by reducing packet protocol overhead through message concatenation algorithms applied along the data collection tree. Such algorithms and policies, however, do not exist currently and need to be developed, keeping in mind the unique characteristics of metering data like variable packet sizes, stochastic arrivals, and the presence of messages with and without deadlines. Current DCUs on the market lack the ability to reduce the volume of data flowing through them and real-time aggregation capabilities. They only provide simple integration of sensing and WAN communications options with the intention of following
the PRIME standard [73, 84] which gives the utilities the freedom to choose meters from various vendors and avoid being reliant on proprietary solutions from a single source.

Figure 4.1: Data concentrator unit’s envisioned role of message concatenation at power distribution level.

In this part of the dissertation, a suite of online message concatenation algorithms at DCUs in the AMI scenario are designed and comparatively evaluated to minimize usage of network capacity in transporting data through the meter data collection network while meeting quality-of-service constraints imposed by applications on individual messages. The specific contributions of this work include the following:

1. A formulation of the message concatenation problem at DCUs in smart metering networks to minimize network capacity utilization.

2. Hardness results for the formulated message concatenation problem that proves it as non-deterministic polynomial(NP)-complete.

3. Six different heuristic-based algorithms that can be employed at DCUs for the message concatenation problem.

5. Exploration of feasibility of message concatenation under practical settings considering network and processing delays, tighter application deadlines, and lossy backhaul links.

Results indicate that the proposed heuristic-based concatenation algorithms can reduce data volume in the range of 10-25% for typical backhaul technologies used, with greater benefits seen for scenarios with higher data traffic rates. These benefits are obtained operating only on packet headers without compressing or aggregating the underlying information in messages. These results are also shown to hold up well under various practical issues such as network and processing delays, tighter application deadlines, and lossy backhaul links.

4.2 Related Work

Numerous prior work has been done on data aggregation in the field of wireless sensor networks, along with a nice survey [75]. Typical approaches in WSNs have focused on efficient data gathering and energy-latency tradeoffs [11, 89]. Having utilized the graph theory [37] Habib and Marimuthu have proposed a model to aggregate sensors data at the gateways within WSNs. In their model, all sensors’ tasks are considered as a directed acyclic graph (DAG), and then the collected data have been grouped into a super task flow graph (STFG). This resulted in smoothly aggregated (scheduled) collected data at the gateway without losing data or overlapping between data. In addition to WSNs, Other studies have designed a reliable, flexible, and cost-effective data concentrator that can connect both a broadband interactive network and a home network. Chen et al. [21] have presented a hybrid channel data concentrator for reliable Internet access control composed of power line communication, (PLC) and wireless front-end modules, and an embedded system. A quantitative channel reliability measure is proposed for channel selection and additional purposes like remote diagnosis and system service improvement. In another work [77], overvoltage protection of data concentrators used in smart grid applications has been studied. Further studies
have even extended the problem of data concentration to synchrophasors. The latest activities discussed are the standardization of wide area monitoring, protection, and control (WAMPAC) systems, and design and implementation issues, such as the time synchronization clock lost at the phasor measurement unit (PMU), missing phasor data frames, handling multiple input data rates and latency from PMUs, etc. with data concentrators are discussed in [3, 91].

However the most related work compared to the work in this dissertation is that of Zhu et al. [92], who have presented a power-efficient scheme to deliver real-time data packets in sensor networks. With this application, each packet is associated with end-to-end deadlines, within which they must reach their destination at a base station (BS). They proposed a state-of-the-art scheme that performs load-balanced routing and distributes data packets evenly among nodes relaying data towards the BS, avoiding bottlenecks and increasing the likelihood that packets will meet their deadlines. They proposed a method of grouping smaller packets into larger ones by delaying data transmissions at the relaying nodes whenever slack times are positive. They claimed that the scheme they have utilized for grouping packets has significantly reduced packet transmissions and congestion and also saved power in the sensor network. The goal in this dissertation is similar— proposing optimal message grouping and packet scheduling at DCUs to reduce data capacity requirements of the communication backhaul. However, power or energy consumption of the nodes employed are not considered because the AMI infrastructure is expected to have access to electric power with backup batteries all the time. This shifts the focus of the problem to the reduction of network capacity utilization, keeping latency constraints in mind with no attention paid to battery life of the nodes involved and the construction of an energy-efficient data-gathering tree. Allalouf et al. [5] also does look at data volume reduction in smart metering networks, but they do not include aspects such as message concatenation. On the other hand this proposed work is to design data concentration algorithms specifically for smart metering and reduce information volume through the network.
4.3 Problem Statement

4.3.1 Motivation

In most communication protocol suites, such as the transmission control protocol/internet protocol (TCP/IP) used for sending smart metering messages, the small size of packets will result in a high amount of protocol overhead due to packet headers. For example, for 100-bytes messages from the source smart meter, there may be 40-60 bytes of additional header overhead due to the TCP/IP protocols and specific versions used. If a data concentrator collects multiple packets and strips off all individual headers and includes only one header for the larger aggregated message, then there could be significant reductions in network capacity utilization. Studying the messaging format, as shown in Figure 4.2, for the ANSI C12 smart meter communications standard provides an idea of message sizes involved and the amount of protocol overhead to expect. Each smart meter-generated message includes parameters such as meter identification number, equipment status, and type of message, among others. This information is enough to uniquely identify a message source with no additional protocol header information required for source identification. Thus, source protocol headers can be stripped away in order to rely only on a common aggregated packet header to route the packet to the destination.

![Smart meter datagram structure.](image)

Table 4.1 (abstracted from the work of Luan et al.[62]) lists the basic types of message along with their properties.
It can be seen that messages can vary in size, have loose or strict deadlines, or have no deadlines at all. Some messages may be generated randomly at any time to indicate critical events that need to be responded to immediately. Data concentrators will have the challenge of handling these varying message sizes, which may or may not have deadlines, with possibly stochastic arrivals, and at the same time guaranteeing that each message meet any specified deadline. Stochastic message generation and critical events with short deadlines exclude the use of polling-based algorithms to collect data at DCUs.

### 4.3.2 Smart Metering Message Concatenation Problem

The smart metering message concatenation (SMMC) problem considered in this dissertation is as follows: A DCU receives different types of messages from smart meters with a stochastic arrival process (arrival process discussed later in Section 5.5). Each message can be of a different size and comes with an application-specific end-to-end deadline by which it must reach the common destination, which is the utility control center. Each message has protocol overhead because it is packaged into a packet before being sent to the DCU. The DCU can either send each packet to the destination as it arrives as a single message or wait and concatenate multiple messages before
sending them out over the backhaul to the destination. The objective considered is to minimize the number of individual packets (and hence protocol overhead) sent upstream by the DCU so as to reduce network capacity requirements of the backhaul. The constraints are that all packets meet their deadline (if any) and that each concatenated packet generated (including a common packet header) has an upper size limit, $W$, governed by the maximum transmission unit (MTU) of the upstream link from the DCU. The objective function chosen helps reduce the total overhead required to send all messages within a given time period $T$ by maximizing the size of each concatenated packet for a fixed header size $H$. In this work, it is assumed that messages are not compressed from their original sizes (zero-compression) and that the solution to the SMMC problem at DCUs would serve as a lower bound for the possible reduction in network utilization by additional schemes (possibly that compress message sizes themselves) developed in the future for the smart metering scenario. In this work, the focus is on only a single DCU and its concatenation operation; in the next chapter, wider view of the backhaul network and the use of multi-level DCUs along the communications network are envisioned.

A formal statement of the SMMC problem is provided in the following definition:

**Definition 1.** Assume that over some period of time $T$, all smart meters together generate $n$ messages $M = \{m_1, \ldots, m_n\}$. Each message $m_i \in M$ has size $s_i$ and an associated protocol header $h_i$ accompanying it until the DCU with $(s_i, h_i, s_i + h_i \in [0, W])$, an arrival time at the DCU of $a_i$ ($a_i \in [0, T]$), and a deadline $d_i$ ($d_i \in [a_i, \infty]$) by which it must leave the DCU, where $i = 1 \cdots n$. Then, the SMMC problem is to determine an integer number of packets $k (k \leq n)$ and a $k$-partition $P_1 \cup P_2 \cup \cdots \cup P_k$ of the set $M$ such that (i) $\sum_{i \in P_j} s_i + H \leq W$, $\forall j = 1 \cdots k$, and (ii) each message $m_i \in M$ meets its deadline with $\max_{i \in P_j} a_i \leq \min_{i \in P_j} d_i$. A solution is optimal if it has minimal $k$.

The SMMC problem can also be stated as a $0 - 1$ integer linear program (ILP) as follows:

$$\text{minimize } k = \sum_{i=1}^{n} y_i$$

subject to constraints
\[
\sum_{j=1}^{n} s_j x_{ij} + H \leq W y_i, \quad \forall i \in \{1 \cdots n\}
\]

\[
\max a_j x_{ij} \leq \min d_j x_{ij}, \quad \forall i \in \{1 \cdots n\}, j \in \{1 \cdots n\}
\]

\[
\sum_{i=1}^{n} x_{ij} = 1, \forall j \in \{1 \cdots n\}
\]

\[
y_i \in \{0, 1\}, \forall i \in \{1 \cdots n\}
\]

\[
x_{ij} \in \{0, 1\}, \forall i \in \{1 \cdots n\}, \forall j \in \{1 \cdots n\}
\]

where \(y_i = 1\) if packet \(i\) is used, and \(x_{ij} = 1\) if message \(j\) is put into packet \(i\).

In the formulations above, the term deadline refers to the local deadline for a message at the DCU by which a particular message must be picked up for the packet creation and transmission over the network. This local deadline can be set by subtracting away an estimate of processing delay at the DCU and the network delay over the backhaul from the end-to-end deadline specification of an application for messages. The impact of processing and network delays will be discussed and incorporated later in Section 4.8. In the problem definition above, for any set of messages assigned to a packet, none of the messages in the packet will miss their local deadlines at the DCU if the arrival times of all messages are at least some value \(\epsilon\) before the first expiring deadline value among all messages of that set. This value \(\epsilon\) could be set to the maximum processing delay to be encountered at the DCU in forming a packet and could be an input to the problem; more discussion about estimation of processing delays will be presented in Section 4.8.

4.4 Algorithms for SMMC Problem

4.4.1 Nature of SMMC Problem

The SMMC problem as stated at the end of the previous section falls into the class of online scheduling problems. Because an offline SMMC problem (with known message arrival times) can be easily reduced to the classical bin-packing problem that is known to be NP-complete
[25], the online version is also NP-complete.\textsuperscript{6} Thus, in this work, heuristic algorithms for solving the SMMC problem are developed. The proposed heuristic solution approach is to rely on earliest deadline first (EDF) scheduling, where a concatenated packet is created at the DCU starting with a message within a specific threshold of its deadline and then filled with other messages so as to maximize the packet size that can be sent out. Proposed heuristic algorithms differ in terms of what other messages they decide to fill in the concatenated packet in addition to the message whose deadline is about to expire. Figure 4.3 depicts the idea described above.

![Figure 4.3: Data concentrator schematic concept.](image)

### 4.4.2 SMMC Hardness Result

To prove that the SMMC problem is NP-complete, first it is shown that SMMC is in NP, in other words, has a polynomial time verifier. An instance of a solution to the SMMC problem is an integer number of packets $k$ and a feasible $k$-partition $P_1 \cup P_2 \cup \cdots \cup P_k$ of the set of messages $M$. Such an instance can be verified in polynomial time in terms of the input consisting of the following fields $<$message identifier, arrival time, deadline, message

\textsuperscript{6}A formal proof is omitted due to space limitations. For interested readers, a similar problem is the scheduling with release times and deadlines on a minimum number of machines (SRDM) [22].
size, header size, $W > n$ messages. Further, in polynomial time (in terms of input length), it is possible to that each message falls in exactly one of the $k$ partitions/packets, and that each packet meets the condition of having its total size less than or equal to $W$. It can be further checked in polynomial time to see if any message in the packet will miss its local deadline. Thus, it can verified whether a given instance is a solution to SMMC in polynomial time and, hence, $\text{SMMC} \in \text{NP}$.

To prove that the SMMC problem is NP-hard, the known NP-complete bin packing-problem [25] is reduced to the SMMC problem. These problems have many similarities but differ in terms of the notion of arrival times and deadlines for the SMMC problem. The bin packing problem takes as input a set of $n'$ items $I = \{i_{t_1}, \ldots, i_{t_{n'}}\}$ of sizes $S' = \{s'_{t_1}, s'_{t_2}, \ldots, s'_{t_{n'}}\}$ and a set of bins $B = \{b_1, \ldots, b_{k'}\}$ each of size $W'$. An assignment of items to bins that minimizes the number of bins $k'$ into which all items are packed is sought. That is, a $k'$-partition $B_1 \cup B_2 \cup \cdots \cup B_{k'}$ of the set of items $I$ is sought.

An instance of the Bin packing-problem to that of the SMMC problem is transformed as follow. For each item $i$ in $I$, add dummy variables $A' : a'_i = 0$, and $D' : d'_i = \infty$. This transformation can be trivially done in polynomial time (in terms of input length) and the modified instance used as an input to the SMMC problem with $M = I$, $S = S'$, $D = D'$, $A = A'$, $W = W'$, and $P = B$.

Any resulting solution from the SMMC problem can be transformed back to a solution for the bin packing problem as follows: A solution to the SMMC problem gives an integer $k$ and a $k$-partition of $M$ that maps individual messages to specific concatenated packets. This solution and can be taken and applied to apply the following transformation: $k' = k$ and $B_i = P_{i-1}$, $i = 1 \cdots k$. This transformation gives the required solution assignment for the bin-packing problem and can be easily done in polynomial time again.

**Theorem 1.** SMMC is NP-complete.

**Proof.** By transforming (in polynomial time) any input instance of the bin-packing problem to that of an SMMC problem, and the resulting solution of the SMMC problem back to Bin Packing
problem, bin-packing has been reduced to SMMC. Thus, SMMC is an NP-hard problem. And since we had proved SMMC $\in$ NP earlier, it can be concluded that SMMC is NP-complete.

The problem as stated so far is an offline version where all packet arrival times and deadlines are known beforehand, and the DCU needs to solve the problem looking forward at the entire window of messages that could arrive over duration $T$. This problem can occur in practice when all message types and their arrival times are known deterministically, for example, when all messages are scheduled deterministically. However, in most cases, the problem will be an online one with stochastic types of message and arrivals, where the DCU will only have access to those messages (with their arrival time and deadlines) that have reached the DCU and are waiting to be concatenated before being sent out over the backhaul. Thus, any proposed heuristics will need to be performed in an online fashion.

4.4.3 Heuristic Algorithms

Due to the proven hardness of the SMMC problem, online heuristic-based algorithms are developed in this work for solving the SMMC problem. The proposed heuristic solution approach is to rely on earliest-deadline-first scheduling, where a concatenated packet is created at the DCU starting with a message within a specific threshold of its deadline and then filled with other messages so as to maximize the packet size that can be transmitted. The proposed heuristics differ in terms of what other messages they decide to fill in the concatenated packet, in addition to the message whose deadline is about to expire.

Six different heuristic algorithms are proposed for scheduling of messages at a DCU for the SMMC problem, as listed in Table 4.2. All six algorithms initiate creating a packet when one of the message deadlines is about to expire; they differ in terms of what other messages (in addition to the message whose deadline is about to expire) are put in the packet being sent out.

In all six schemes, the aggregation process starts with the main procedure, which is required to be run consistently over the DCU. The data concentrator unit-optimized message allocation (DCU-OMA) procedure is an infinite loop as defined in the main Procedure 1, which follows.
Table 4.2: PROPOSED CONCATENATION HEURISTICS

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EDF-DKB</strong></td>
<td>Inserts deadline messages as much as possible inside the packet, and the remaining space is filled through knapsack selection over best-effort messages that have been queued.</td>
</tr>
<tr>
<td><strong>EDF-SDKB</strong></td>
<td>Only a single-deadline message sits inside the packet, with any available space filled with non-deadline messages in the non-deadline queue through knapsack selection.</td>
</tr>
<tr>
<td><strong>EDF-FCFS</strong></td>
<td>Messages will be placed in the packet according to their arrival sequence from a common queue of deadline and non-deadline messages on a first-come first-served basis.</td>
</tr>
<tr>
<td><strong>EDF-KN</strong></td>
<td>Messages are chosen from a common pool of deadline and best-effort messages selected through the knapsack algorithm.</td>
</tr>
<tr>
<td><strong>EDF-KDKB</strong></td>
<td>A sequence of knapsack selections, working first on all queued deadline messages and then on queued best-effort messages, if needed to fill the packet.</td>
</tr>
<tr>
<td><strong>EDF-KBKD</strong></td>
<td>Reverse order of the knapsack process in EDF-KDKB, working first on the queued best-effort messages and then on the deadline messages, if needed.</td>
</tr>
</tbody>
</table>

The “Classifier” module checks the arrived messages to see whether they are best-effort or have a specific deadline (if the selected heuristic requires to differentiate between them). Two different queues observable in Figure 4.3 are formed based on the classification done. All deadline messages are kept in a priority queue sorted by earliest deadline. It is assumed there are two queues in the system: one for the messages with a specific delay objective \( Q_d \) and the other for those without a delay objective (best-effort messages, \( Q_{nd} \)). The call to function `bufferQueues` inserts the messages with deadline to the first queue using sub-function `enqueue_d(M)` and buffers the best-effort messages into the second queue using sub-function `enqueue_nd(M)`. If no classification is required, then all arrived messages will be sorted and placed in a single buffer using `enqueue(M)`.

The “Scheduler” module always checks the queues and does the appropriate operation when they contain messages. Here, the EDF algorithm is utilized.

\( \tau \) is a threshold value that the DCU will not send packets with sizes less than that unless there are no other messages or a deadline is passing.
**Procedure 1 Main**

1: procedure DCU-OMA
2: alg← selectAlgorithm()
3: repeat                       \(\triangleright\) forever
4:    bufferQueues(alg)
5: until DCU fails.
6: end procedure

**Module 1 - Classifier**

8: function BUFFERQUEUES
9: if alg is “EDF-FCFS” or “EDF-KN” then
10:    enqueue(M)
11: else
12:    if delay objective is best-effort then
13:        enqueueM(M)
14:    else
15:        enqueueD(M)
16: end if
17: end if
18: end function

**Module 2 - Scheduler**

19: function CHECKQUEUES
20: while Queue(s) are not empty do                     \(\triangleright\) Drop messages that have missed their deadline
21:     drop(Mmissed)
22:     EDFsort
23:     pkt ← createPacket()
24:     sendPacket(pkt)
25: if (addNDM.size > \(\tau\)) then\(^7\)
26:     addNDM()
27:     sendPacket(pkt)
28: end if
29: end while
30: end function

A packet is formed “just-in-time” before the deadline of the first message in the priority queue expires,\(^8\) with each of the heuristic algorithms having a different approach on how to maximize the size of the packet that is sent out.

During the main procedure the most important call is to the createPacket procedure. Assume \(M_i\) is the \(i^{th}\) message in the deadline queue. The function dequeue\((M_i)\) moves message \(M_i\) from the queue of DCU to the outgoing packet, whereas function requeue\((M_i)\) inserts the message \(M_i\) back to the end of the queue, if it was not selected during knapsack selection. Finally, it is assumed that at any given time TPS, is the current “total packet size” inside the DCU.

---

\(^7\)Even though the deadline for a message is the time by which it must reach its eventual destination over the network from the DCU, we assume the value of deadline can be shifted by some constant \(T\), where the DCU must be fairly confident that most packets will suffer a delay \(< T\) over the backhaul network. Future work could look at best ways to estimate \(T\); some known approaches are to take use the 95% value over a historical window of latencies or use a weighted sliding window as used in [71].
Procedure 2 Packet Creation

1: procedure CREATE_PACKET
2: TPS ← 0
3: i ← 1
4: deadline ← currentTime + Ti
5: switch alg do
6: case 1 - EDF-DKB
7: dequeue(Mi(Qd))▷ Put the first messages with deadline in the packet
8: while size(TPS + Mi + Header) ≤ MTU do
9: if (deadline is not passed) then
10: dequeue(Mi)
11: else
12: requeue(Mi)
13: end if
14: end while
15: addNDM()
6: case 2 - EDF-SDKB
16: Do the same procedure of EDF-DKB but with a single loop
6: case 3 - EDF-FCFS
17: while size(TPS + Mi + Header) ≤ MTU do
18: dequeue(Mi)
19: TPS ← append(TPS, Mi)
20: end while
6: case 4 - EDF-KN
21: while (deadline is not passed) do
22: waitList ← dequeue(Mi)
23: updatedeadline
24: end while
25: TPS ← knapsack0−1(waitList)
6: case 5 - EDF-KDKB
26: while (deadline is not passed) do
27: waitList ← dequeue(Mi(Qd))
28: updatedeadline
29: end while
30: TPSd ← knapsack0−1(waitList)
31: if size(TPS + Header) ≤ MTU then
32: TPSnd ← knapsack0−1(Qnd)
33: end if
34: TPS ← append(TPSd, TPSnd)
6: case 6 - EDF-KBKD
35: dequeue(Mi(Qd))▷ Put the first messages with deadline in the packet
36: while (deadline is not passed) do
37: waitList ← dequeue(Mi(Qnd))
38: updatedeadline
39: end while
40: TPSnd ← knapsack0−1(waitList)
41: if size(TPSnd + Header) ≤ MTU then
42: TPSd ← knapsack0−1(Qnd)
43: end if
44: TPS ← append(TPSd, TPSnd)
45: return append(TPS, Header)
46: end procedure
That is, the packet is being formed using the packet creation, Procedure 2, which, when called, will create the packet depending on the heuristic algorithm selected.

Since messages with the delay objective always have the higher priority to the best-effort one, it is important to know when and how to add non-delay objective messages to the outgoing packet. The best-effort message addition, Procedure 3, utilizes the 0-1knapsack algorithm [25] to accurately append best-effort messages to the packet when possible.

**Procedure 3 Best-Effort Messages Addition**

1: \textbf{procedure} \texttt{ADDNDM} \hfill \triangleright Add Non-Deadline Messages to Packet
2: \hspace{1em} \texttt{B[i] } \leftarrow \textbf{benefitFP}(M_i(Q_{nd}))
3: \hspace{1em} \texttt{TPS}_{nd} = \textbf{knapsack}_{0-1}(B[i], Q_{nd})
4: \hspace{1em} \texttt{append}(\texttt{TPS}, \texttt{TPS}_{nd})
5: \hspace{1em} \texttt{return} \texttt{TPS}
6: \textbf{end procedure}

The function of \textit{benefitFP}(N) in Procedure 3 is to check the resulting benefit out of placement of all messages in a waiting list during packet creation and stores the messages in an array of positive integer values, \texttt{B[i]} (the inverse value of remaining space after placement of a particular message in packet). This array is then used in the 0-1knapsack algorithm.

### 4.5 Reference Algorithms

#### 4.5.1 EDF-Based Integer Linear Programming (ILP) Formulation

The SMMC problem can be solved using mathematical optimization algorithms. The SMMC problem has been formulated as a mixed-integer linear program, which optimally schedules the remaining messages in addition to the EDF message to begin a packet with an index. For a packet with index \textit{i}, the problem is formulated as follows:

\[
\begin{align*}
\text{maximize} \quad & P_i = \sum_{j=1}^{n_i} s_j x_{ij} \\
\text{subject to constraints}
\end{align*}
\]
where \( x_{ij} = 1 \) if message \( j \) is put into packet \( i \). In the formulation above, \( n_t \) \((n_t \leq n)\) is the set of messages queued at the DCU and available for concatenation at time \( t \) \((t \leq T)\). Any messages that are found to not meet deadline constraints are forwarded immediately with no concatenation process applied. This formulation is different from equation (5.1) in that it is EDF-based and message deadlines are not a constraint, because messages closest to their deadlines are selected and sent out before their deadlines occur. This formulation tries to fit in as many messages as possible (among those available) in a packet to be sent out. The given constraint specifies the maximum packet size that can be sent over the backhaul technology with a specific MTU size. The drawback of this approach in practice (as opposed to the heuristics) is the brute force nature of this integer linear programming solution procedure, which makes it practically infeasible for real-time applications and those that involve large-scale data.

### 4.5.2 Theoretical Optimal

This method is theoretically the minimal number of packets that needs to be sent from a DCU for a given number of messages generated from the smart meters over a period of time. This value is not constrained by arrival times or message deadlines of messages; it is computed over all generated messages and maximum packet size MTU that includes a header size \( H \). This value can be mathematically computed by:

\[
\min(Num_{Packets}) = \left\lceil \frac{\sum_{i=1}^{n} s_i}{MTU - Header} \right\rceil
\]

where \( n \) is the total number of arrived messages during a time interval, \( s_i \) is the size of a message \( i \), and MTU and header size \( H \) are the parameters defined according to the backhaul technology.
Although this solution is not feasible in practice, it gives a theoretical reference for the performance evaluation of any SMMC algorithms, not limited, EDF-based heuristics.

4.6 Evaluation

4.6.1 Methodology

Outlined below are more details about the simulation environment, message arrival process, and distribution of various types of messages.

4.6.2 Simulation Environment

A discrete-event simulator was developed using MATLAB to evaluate the proposed heuristic algorithms and compare them to the reference algorithms. For the EDF-based ILP algorithm, the “bintprog” function from the optimization toolbox was used as the solver. The network topology consisted of a group of smart meters generating messages as a Poisson arrival process and sending messages to the DCU to be routed to the control center.\textsuperscript{9} Due to the assumption of each meter generation as a poisson process, we can sum individual average message generation rates to get an average arrival rate at the DCU of $\lambda$ which is used as a parameter. We have considered three different $\lambda$ values of 0.1, 0.5, and 1 at the DCU which would correspond to 90, 450, and 900 smart meters sending 1 message on average every 15 minutes. The service capacity of the DCU is considered to be infinite. An assumption made in this work is that processing delays at the DCU can be ignored in constructing a concatenated packet from a queue of messages.\textsuperscript{10} The simulation time interval was kept at 300 seconds.

\textsuperscript{9}Prior work [57] supports this assumption that smart meters message generation can be modeled as a Poisson process.

\textsuperscript{10}This assumption would not make a difference in comparing all proposed heuristics, if queue sizes remain relatively small; however, for large queue sizes, there may be an impact that does not come out in the evaluation results, as explained in the following section.
4.6.3 Distribution of Message Types

During a day, different types of messages may be exchanged between smart meters and the utility control center through the AMI. In these evaluations, all seven basic types of messages listed previously in Table 4.1 and first reported by Luan et al. [62] are considered. Based on geographic location, power distribution infrastructure, and utility preferences, the transmission of messages could come from different distributions of these basic message types, which will have an impact on the performance of the proposed heuristics. In these evaluations, different Beta distributions across these message types were used by varying the shape parameters, $\alpha > 0$ and $\beta > 0$.

Assuming that the arrival probability of different message types is $p_1, p_2, \ldots, p_n$ with $0 < p_i \leq 1$ for $i = 1 \cdots n$ and $\sum_{i=1}^{n} p_i = 1$, each can be computed as the result of the difference between cumulative probabilities of $n$ intervals as:

$$p_i = \text{betacdf}(i) - \text{betacdf}(i - 1)$$

where $\text{betacdf}$ is the cumulative density function of the beta distribution used with parameters $\alpha$ and $\beta$.

For the experiments, five different message-type distributions were generated using the shape parameters shown in Table 4.3 to test the performance of the proposed algorithms.

4.7 Simulation Results

One hundred runs of simulations were conducted with mean values along with 95% confidence intervals plotted in the results. Each scheme was evaluated in terms of the overall reduction in bytes of data transmitted out into the backhaul network by the DCU, compared to the overall incoming data in bytes from smart meters, including all headers. Each packet header was assumed to be a fixed size of 50 bytes corresponding to the 40-60-byte range for TCP and IP headers. Figure 4.4 displays the output of the proposed algorithms and reference algorithms over five message-types distributions with 95% confidence intervals.
Table 4.3: PRE-DEFINED MESSAGE ARRIVAL DISTRIBUTIONS

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform ($\alpha = 1, \beta = 1$)</td>
<td>The message traffic has an almost equal percentage of all message types.</td>
</tr>
<tr>
<td>More smaller ($\alpha = 2.8, \beta = 1.9$)</td>
<td>Most of the arrived messages are smaller-sized messages.</td>
</tr>
<tr>
<td>More larger ($\alpha = 0.18, \beta = 0.25$)</td>
<td>Most of the arrived messages are larger-sized messages, with very few smaller-sized messages.</td>
</tr>
<tr>
<td>More deadline ($\alpha = 1, \beta = 1.8$)</td>
<td>Most of the arrived messages have a deadline restriction.</td>
</tr>
<tr>
<td>More best-effort ($\alpha = 2.5, \beta = 0.5$)</td>
<td>Few of the messages have a deadline, so many of these are best-effort messages.</td>
</tr>
</tbody>
</table>

Results are shown for packet arrival rates at the DCU of $\lambda = 0.1$, 0.5, and 1. It can be seen that the overall data volume reduction varies 5-25%, depending on the message-type distribution, message arrival rate at DCU, and specific algorithm used. From these results, three questions are answered: (1) How do the proposed heuristic algorithms stack up against each other and against the reference algorithms? (2) What is the impact of message-type distribution? and (3) What is the impact of $\lambda$?

4.7.1 Effect of Proposed Heuristic Algorithms

The bar charts in Figure 4.4 show that the EDF-KN algorithm has the best performance among all other heuristic algorithms and comes very close to the performance of the MILP across all $\lambda$ and message type distributions. This is because EDF-KN is using a common pool of messages, whether they be deadline or best-effort, thus providing more options to maximize the packet size before it is sent out. Since typically there are enough queued messages before a deadline is reached, the algorithm has a good collection of options to maximize the packet before sending it out.
4.7.2 Impact of Message-Type Distribution

The uniform distribution of all message types serves as the reference case to compare with other distributions. For the “more deadline” case where the majority of all messages have deadlines, the overall data volume reduction is smaller for all algorithms. The presence of more messages with deadlines than best-effort necessitates that packets be sent out of the DCU without having the luxury of waiting for the right combination to maximize packet size. On the other hand, when there are mostly best-effort packets present, algorithms can wait longer before being forced to send out packets; this allows each packet to be larger, and hence reduces packet overheads. The case for more smaller-sized messages is similar to the more-deadline message case in that it helps reduce packet overhead significantly through concatenation because header sizes are comparable to data sizes. Smaller messages are also easier to group into a packet. Conversely, the more-larger messages case results in greater difficulty to group messages into a packet; also larger underlying message sizes already have reduced overhead making improvement through concatenation difficult.

4.7.3 Impact of $\lambda$

The value of $\lambda$ signifies the packet arrival rate at the DCU; hence, larger values indicate that more messages are arriving at the DCU, and more packets need to be sent out after concatenation. With greater data volume, there are more opportunities for each concatenation algorithm to find a best fit of messages in an outgoing packet from the DCU to reduce overall protocol overhead. The EDF-KN data volume reduction approaches very close to that of even the theoretically optimal solution with increasing $\lambda$. Thus, with a greater rate of packet arrivals, the proposed EDF-based concatenation algorithm over a common queue of messages maximizes the reduction in data volume.
Figure 4.4: Overall data reduction percentage using proposed heuristics over different message arrival rate and message type distributions.
4.8 Impact of Network and Processing Delays

Network delays between the DCU and the utility control center, and processing delay at the DCU itself are two factors that are assumed to be negligible in the results presented so far. The magnitude of these delays may not be negligible in all practical cases, and can cut down the amount of time a DCU can wait to maximize the size of outgoing packets sent out. Thus, there will be a direct correlation between network and processing delays on the ability of a DCU to reduce protocol overhead. An interesting challenge here is that the DCU cannot accurately predict these delays beforehand; each concatenated packet will suffer variable network and processing delays due to many factors related to number of messages processed and characteristics of the communication backhaul. Thus, the DCU must rely on an estimate of network and processing delays needed to budget into computing the local deadline of each message. An overestimate will reduce the amount of time a DCU will have to wait and concatenate a large packet; on the other hand an underestimate can mean some messages will miss their deadlines. This section describes how such delays can be estimated and what impact it will have on data volume reduction through message concatenation.

4.8.1 Estimation of Network and Processing Delays

To estimate the processing delay, it was necessary to examine it into the major individual components that cause delay. These components are as follows: (i) concatenation delay: the time required to put all selected messages into a packet and add a common header; (ii) knapsack delay: the time required by some of schemes that use a knapsack operation to select messages from a queue of messages; and (iii) sorting delay: the time required to maintain the queue, sorted in terms of earlier deadlines. These components are present in each heuristic in possibly different ways based on the nature of the algorithm. Table 4.4 summarizes how each of these components ($C_C$, $C_S$, and $C_K$, which are the time costs for concatenation, sorting, and selection through knapsack, respectively) add up to the total processing delay for each heuristic scheme. These schemes operate
on either a single common queue of \( n \) items, or one of two queues (with sizes \( n_1 \) and \( n_2 \)) having deadline and non-deadline messages, or both queues one after the other.

Table 4.4: HEURISTICS PROCESSING-TIME CALCULATIONS

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Processing Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDF-FCFS</td>
<td>( C_C + C_S )</td>
</tr>
<tr>
<td>EDF-KN</td>
<td>( C_C + C_S + C_K(n) )</td>
</tr>
<tr>
<td>EDF-DKB</td>
<td>( C_C + C_S + C_K(n_1) )</td>
</tr>
<tr>
<td>EDF-SKB</td>
<td>( C_C + C_S + C_K(n_2) )</td>
</tr>
<tr>
<td>EDF-KDKB</td>
<td>( C_C + C_S + C_K(n_1) + C_K(n_2) )</td>
</tr>
<tr>
<td>EDF-KBKD</td>
<td>( C_C + C_S + C_K(n_2) + C_K(n_1) )</td>
</tr>
</tbody>
</table>

The next step was to populate realistic values into the processing delay estimation model. To do this, the actual processing delays were measured when executing each of the three operations: concatenation, knapsack selection, and keeping a sorted queue. These values were computed on a Dell Optiplex 64-bit PC with a 2-core 2.8 GHz CPU and 5 GB RAM for a full range of values of \( n \) from 1 to 1000 to study all possible queue sizes likely to be encountered for message arrival rates used in the evaluations in Section 5.5.\(^\text{11}\) By populating these values for a given \( n \) in the processing delay model presented in Table 4.4, the DCU could easily construct an estimate.\(^\text{12}\)

### 4.8.2 Evaluation Results

Here the proposed heuristic-based algorithms are reevaluated with varying values of network and processing delays, and the impact on achievable reductions in protocol overhead are studied. For these evaluations, the EDF-KN heuristic is chosen, because it is one of the better performing heuristics among those evaluated in Section 5.6. Figure 4.5 presents the results for \( \lambda = 1 \).

\(^\text{11}\) It is assumed that when proposed algorithms are deployed, an estimate can be recalculated for the specific system employed in the DCU as opposed to using the estimates discussed here. DCUs on the market can have high processing capabilities as described in [34], and it is expected that the values used in this work over-estimate the actual processing delays.

\(^\text{12}\) For more details, the reader is referred to the work of Namboodiri and Gao [71] for a description of how network delays can be predicted with an exponentially weighted moving average over a sliding window of previously seen delays. The impact of various possible network delays will be studied to assess the impact on benefits of message concatenation in evaluations that follow.
The bars are showing the protocol overhead reduction achieved with varying values of network and processing delays, including the case where such delays are set to nil. In addition, to further explore the lower limits of possible benefits of message concatenation, the experiments are reperformed with deadline values that are half and a quarter of the amount of that used in the evaluations in Section 5.5. The impact of tighter deadlines will be similar to that of additional network and processing delays, with both factors essentially reducing the time the DCU has to concatenate messages into larger packets.

The results shown in Figure 4.5 indicate that as processing and network delays increase, the percentage of overhead reduction decreases. Similarly, as deadlines get tighter, the achievable data volume reduction is reduced. Even for such extreme cases considered, there is at least a 5% possible reduction in data volume. The biggest impact of network and processing delays, or tighter deadlines, is with the “more deadline” message distribution, with a greater fraction of messages needing to be concatenated and sent out quickly. The smallest impact of delays or tighter deadlines is seen for the “more best-effort” case, where most messages are not hard pressed to meet deadlines.
A more accurate depiction of what happens inside the DCU can be seen by studying the average queue or buffer size for various message-type distributions for estimated processing delays and varying network delay values of 100 ms and 250 ms. A similar trend can be expected for tighter deadline values. As Figure 4.6 confirms, the “more deadline” message distribution has the smallest average queue size, implying that messages do not stay in the buffer for long periods. At the other extreme, the “more best-effort” message distribution results in the largest average queue size, implying that messages stay in the buffer for a much longer duration. A large average queue size does add additional processing delay; however, for the “more best-effort” case, there are few messages with deadlines that are impacted by the larger processing delays. For all the other schemes, as evident from the results, the average queue size stays small enough to not adversely impact data volume reduction.

4.9 Data Volume with Lossy Links

Another practical aspect that needs to be considered is the impact of lossy backhaul links on the large concatenated packets expected to be sent out from the DCU by proposed heuristic-based algorithms. Larger packets will typically suffer more retransmissions (thus adding to the volume of data transported) when sent through networks with a fixed bit-error rate (BER) due to
their larger size. Thus, it is imperative to explore the impact of various backhaul technologies, each with different BER characteristics, on benefits of message concatenation\(^{13}\).

### 4.9.1 Theory

The most important factor in analyzing the impact of lossy networks is considering the BER of the technology being used. The transmission BER is the number of detected bits that are incorrect before error correction, divided by the total number of transferred bits (including redundant error codes). Different communication technologies have different BERs. The goal here is to translate a given BER for a technology and estimate the corresponding data volume reduction ratio. Let \(e_b\) be the BER of a given technology. A packet is declared incorrect if at least one bit is erroneous. Thus, for a packet of size \(L\) bits, the resulting packet error rate (PER) of the technology, \(e_p\), is \(e_p = 1 - (1 - e_b)^L\). Let \(D\) be the volume of data in bytes (including payload and control overhead) that would have been sent over the backhaul in a time period \(T\) when message concatenation is not employed. Let \(D'\) be the volume of data sent (again including payload and control overhead) over the backhaul after message concatenation. With a PER of \(e_p\) and \(e_p'\), respectively, the corresponding data volume sent through the backhaul will be \((1 + e_p)D\) and \((1 + e_p')D'\), respectively. Thus, the data volume reduction ratio \(\rho\) with a lossy backhaul can be computed as

\[
\rho = \frac{(1 + e_p)D - (1 + e_p')D'}{(1 + e_p)D}
\]

With larger packet sizes, \(e_p' > e_p\), thus reducing the data volume reduction ratio as compared to the case when lossiness of the backhaul network is ignored.

### 4.9.2 Numerical Evaluation

The technologies considered for the backhaul were fiber optic, WiMAX, and 3G Cellular. These three technologies are currently commonly used to connect the AMI at the customer end.

\(^{13}\)This work does not explore the analogous issue of packet loss due to network congestion; the eventual impact on the benefits of message concatenation is expected to be similar, regardless of the underlying reason for packet loss.
to the backbone network, and they tend to be lossier than the core network. BER values for these technologies based on known ranges in [50, 78, 79] were chosen to study the impact of message concatenation algorithms. The BER values, $e_b$, used in the following evaluation were 5E-07, 3.16E-06, and 7.5E-06 for fiber optic, WiMAX, and 3G cellular technologies, respectively. For each technology, the PER values $e_p$, were computed using equation (4.3). For the case with no message concatenation, an average packet size of 100 bytes ($L = 800$ bits) was considered in computing a PER of $e_p$; for the case with concatenation, a packet size of 1,000 bytes ($L = 8000$ bits) was used to compute $e_p'$, which is roughly the average size of the concatenated packet seen in simulations from earlier sections. Finally, using equation (4.3), $\rho$ was computed for each of the three technologies, with $D$ and $D'$ computed based on simulations discussed earlier in Section 5.5 for the EDF-KN scheme with a message arrival rate of $\lambda = 1$.

It can be seen from Figure 4.7 that for even the most lossy technology considered (3G) with the worst-case BER characteristics chosen, data volume reduction with message concatenation only falls by 3-4%, compared to the reference ideal BER case. Thus, the benefits of message concatenation seems to hold up for the most commonly used technologies. These results are likely to be better with the use of forward error correction (FEC) techniques employed to minimize packet loss.

![Figure 4.7: Data reduction savings vs. different backhaul technologies.](image-url)
4.10 Conclusion

This part of the dissertation has demonstrated that message concatenation algorithms can be an important element of data concentrators deployed in smart grids to solve the looming challenge of transporting massive data volumes through last-mile bandwidth-constrained backhaul networks. Effective message concatenation algorithms at DCUs (such as the EDF-KN algorithm proposed in this dissertation) were shown to be able to reduce overall data volume by 10-25% for each DCU. This reduction was achieved just by a reduction in protocol overhead with no compression of the original data sent by smart meters. Consequently this leaves much room to develop additional data concentration mechanisms at DCUs. Since this work was published, additional research has been conducted in the area of data aggregation in smart grid and AMI applications.
CHAPTER 5
SQUEEZE: AN EFFICIENT MULTI-LEVEL METER DATA-COLLECTION ALGORITHM FOR SMART GRIDS

5.1 Introduction

The information communication and control layer of the smart grid brings about numerous advances, including the empowerment of customers to actively participate in the maintenance of a supply-demand balance around the clock and reliability improvements in electricity service [23, 32]. Advanced metering infrastructure initiatives are a popular tool to incorporate these changes for modernizing the electricity grid, reducing peak loads, and meeting energy-efficiency targets. AMI uses technology to capture and transmit energy use to a collection point on an hourly or subhourly basis, in contrast to standard meters that provide a daily energy usage total and a cumulative monthly bill [28]. With the introduction of AMI technology, two-way communication between a “smart” meter and the utility control center, as well as between the smart meter and consumer appliances, would be facilitated for demand-response, dynamic pricing, cold-load pick-up, and greenhouse gas-emission mitigation [87]. The AMI can also be used for distributed monitoring of the state of the grid in addition to customer premise operations [41]. There are many environmental and economic benefits to utilities, consumers, and society as a whole from adopting AMI technologies [72].

A big challenge for smart grid application scenarios, and the information-sharing framework that enables them, will be handling the massive amount of data that is expected to be collected from data generators such as smart meters and sent through the communication backhaul to the grid operator. For example, by current standards, each smart meter sends reports every 15-60 minutes to a smart meter [27, 74]. When the numbers of such distributed monitoring entities are scaled up to many thousands, the resulting data volume will stress network capacities, especially in limited-bandwidth last-mile networks that are common in metering scenarios at the power distribution level [5, 62]. Future applications may require data to be collected at a finer granularity, thus adding
to the challenge [29]. Network capacity is a precious resource for electric utilities because they are either leasing such networks from third-party providers [43], or building infrastructure themselves and leasing bandwidth out (especially at the backhaul) to recuperate investment costs [51]. In either case, it is in the interest of electric utilities to reduce the volume of information transported through these networks for smart grid applications while ensuring application QoS requirements are met.

The smart grid communication network creates a new traffic engineering challenge requiring a data volume concentration process in order to transmit flows over the network while reducing infrastructure costs. One approach to reducing data volume given some application sampling rate is to concatenate multiple messages into a larger packet to reduce protocol overhead due to packet headers as shown in Figure 5.1. This approach has the potential to reduce network capacity requirements significantly (quantified later in this chapter) due to the small size of messages sent in smart metering networks, with packet headers possibly being of a comparable size to the underlying message to be sent. Such of message concatenation can be done by each smart meter itself. However, each meter may not generate messages frequently enough to be able to have the chance to concatenate enough packets to reduce overheads significantly and also meet their stated application deadlines.

![Figure 5.1: Message concatenation at data concentrator units to reduce protocol overhead.](image-url)
Each meter is also expected to be relatively constrained (compared to a concentrator) in terms of data storage capabilities to keep a large window of packets from which to aggregate. Thus, a better approach is to concatenate messages at an intermediate point upstream from individual meters.

Such an intermediate point where message concatenation can be done is at data concentrator units (or some similar entity, sometimes also called a data aggregator or data collector) that collect data from many smart meters and forward them upstream. Figure 5.2 depicts this concept and shows the DCU’s role at the power-distribution level of the power grid. Data concentrators can play an important role in reducing network capacity requirements by reducing packet protocol overhead through message concatenation algorithms applied along the data collection tree. Such algorithms and policies, however, do not exist currently and need to be developed, keeping in mind the unique characteristics of metering data like variable packet sizes, stochastic arrivals, and the presence of messages with and without deadlines. Current DCUs on the market lack the ability to reduce the volume of data flowing through them and real-time aggregation capabilities. They only provide simple integration of sensing and WAN communications options with the intention following the PRIME standard [73], which gives the utilities the freedom to choose meters from various vendors and avoid being reliant on proprietary solutions from a single source.

Figure 5.2: Multi-level data concentration scenario.
In this dissertation, a suite of online message concatenation algorithms designed and comparatively evaluated at multi-level DCU scenarios in the AMI scenario that efficiently transport meter data through the data collection network while meeting QoS constraints imposed by applications on individual messages. The specific contributions of this work include the following:

1. A formulation of the message concatenation problem in multi-level DCUs in smart metering networks to minimize data transportation costs.

2. Heuristic-based algorithms that can be employed at DCUs for the message concatenation problem.


Results indicate that the proposed heuristic-based concatenation algorithms help minimize costs to transport data over the communications network in the range of 2-30%, with greater benefits seen for scenarios with higher data traffic rates. These benefits are obtained by operating only on packet headers without compressing or aggregating the underlying information in messages. Results are also shown to hold up well under various practical issues such as network and processing delays, tighter application deadlines, and lossy backhaul links.

5.2 Background and Related Work

There has been much prior work on data aggregation in the field of wireless sensor networks [75]. Typical approaches to WSNs have focused on efficient data gathering and energy-latency tradeoffs under deadline constraints (e.g. [11, 37, 38, 39, 89, 92]). These schemes propose algorithms for grouping smaller packets into larger ones by delaying data transmissions at the relaying nodes whenever slack times are positive, with significant reductions in packet transmissions, congestion, and battery energy use. The goal in this project is similar to proposing data concentration at DCUs as relay nodes. However, power or energy consumption of the nodes employed are
not considered because the AMI infrastructure is expected to have access to electric power with backup batteries at all times. This shifts the focus of the problem from the battery life of nodes involved to the reduction of network capacity utilization. The work Allalouf et al. [5] examines data volume reduction in smart metering networks, but does not include aspects such as message concatenation and the application of aggregation functions.

Another direction of related work has been in terms of designing a reliable, flexible, and cost-effective data concentrator [21, 77]. Many studies have considered the problem of data concentration for synchrophasors ([91]). The latest activities in standardization of WAMPAC systems, and design and implementation issues, such as maintaining time synchronization at PMUs, missing phasor data frames, handling multiple input data rates and latency from PMUs, etc. with data concentrators are discussed in the work of Adamiak et al. [3]. On the other hand, the proposed work is to design data concentration algorithms specifically for smart metering and reduce information volume through the network.

Scheduling under deadlines poses well-known challenging problems with many new applications. It was shown by Karp [48] that optimal offline scheduling for the problem of deadline scheduling is NP-complete. On the other hand, simple online scheduling algorithms that achieve the best competitive ratio do exist. For example, the EDF algorithm works on the job with the earliest deadline, and it switches to a newly arrived job if the new arrival has an earlier deadline. It is known that such a simple scheduling algorithm is optimal when the traffic load is light. See in particular the seminal work of Liu and Layland [59], Mok [67], Locke [61], recent applications in scheduling jobs for cloud systems [19], and large-scale electric vehicle charging [20].

Finally, from an information needs perspective, there has been recent work on a futuristic approach to information-sharing mechanisms in smart grids, including those at the distribution level [14, 35, 60, 65, 76]. The GridStat effort [9, 40, 44], primarily from Washington State University, has set about defining communication requirements for power grids for the last 5-10 years. GridStat has further inspired the NASPINet effort to develop an "industrial grade," secure, standardized, distributed, and expandable data communications infrastructure to support synchropha-
sor applications in North America [69]. None of the prior work in the area has looked at information needs of grid operators from a purely information volume perspective and its impact on the design of a communications network for the distribution system.

5.3 Design Objectives and Problem Statement

5.3.1 Motivation for Message Concatenation

In most communication protocol suites (e.g., TCP/IP) used for sending smart metering messages, the small size of packets will result in a high amount of protocol overhead due to packet headers. For example, for messages of size 100 bytes from the source smart meter, there may be 40-60 bytes of additional header overheads due to TCP/IP protocols and specific versions used. If a data concentrator collects multiple packets and strips off all individual headers and includes only one header for the larger aggregated message, there could be significant reductions in network capacity utilization. Studying the messaging format for the ANSI C12 smart meter communications standard in [80] provides an idea of message sizes involved and the amount of protocol overhead to expect. Each smart meter-generated message includes parameters like-meter identification number, equipment status, and type of message, among others. This information is enough to uniquely identify a message source with no additional protocol header information required for source identification. Thus, source protocol headers can be stripped away to rely only on a common aggregated packet header to route the packet to the destination.

As shown previously in Table 4.1 (abstracted from the work of Luan et al.[62]) basic types of message can vary in size, and have loose or strict deadlines, or have no deadlines at all. Some messages may be generated randomly at any time to indicate critical events that need to be responded to immediately. Data concentrators will have the challenge of handling these varying message sizes, which may or may not have deadlines, with possibly stochastic arrivals, and at the same time guaranteeing that each message meet any specified deadline. Stochastic message generation and critical events with short deadlines exclude the use of polling-based algorithms to collect data at DCUs.
5.3.2 Network Topology

This work assumes a multi-level topology of DCUs collecting data from smart meters and transporting them to an electric utility control center. The set of DCUs (apart from the control center and smart meters) is assumed to be organized as a $b$-ary tree with the root DCU node connected to the control center and groups of smart meters connected to leaf DCU nodes. Such a topology with $b = 2$ is shown in the multi-level data aggregation (MLDA) in Figure 5.3.

![Figure 5.3: MLDA logical DCU 2-ary tree.](image)

The multi-level smart metering message concatenation (MSMMC) problem considered in this chapter is as follows: Each leaf DCU receives as packets different types of messages from smart meters connected to it with a stochastic arrival process (discussed later in Section 5.5). Each packet can be of a different size and comes with an application-specific end-to-end deadline by which it must reach the utility control center (that is connected to the root DCU node). A DCU can either forward each packet to its parent DCU as it arrives as a packet, or concatenate multiple messages within received packets into bigger packets with a common header before forwarding to its parent.
To consider the tradeoffs involving data volume, it is assumed that there is a cost of \( c_l \) units per byte of data transported through each link between DCUs at each level \( l \) in the tree. The objective then is to minimize the overall cost \( C \) to transport all data (sum of messages and overhead) generated at smart meters within some period \( T \) to the control center. Because the focus of this work is on reduction in protocol overhead, the overall cost to transport all data depends only on the effectiveness of the concatenation process. The first constraint is that all messages meet their deadlines (if any), and that each concatenated packet generated (including a the packet header) at a DCU at level \( l \) has an upper size limit, \( W_l \), governed by the maximum transmission unit (MTU) of the upstream links of DCUs at that level. It is assumed that the MTU size is non-decreasing as one goes up the tree to ensure concatenated packets need not be fragmented. A study of common MTU sizes for network topologies supports this assumption of non-decreasing MTU sizes from last-mile networks to backbone networks \[33\]. The second constraint is that each packet is formed at a DCU after all its constituent messages have arrived at the DCU and before the first expiring deadline of its constituent messages at the control center, which can be elaborated as follows: At each DCU \( i \), \( k \) packets arrive and \( r \) packets leave \( r \leq k \). Each constructed packet \( j \) \((j = 1 \cdot \cdots r)\) at DCU \( i \) has a deadline \( d_{ij} = \min_{k \in P_{ij}} d'_{ik} \) where \( d'_{ik} \) is the deadline of packet \( k \) as it arrives at the DCU. Each packet constructed at DCU \( i \) cannot be before \( \min_{k \in P_{ij}} a_{ik} \), where \( a_{ik} \) is the time that packet \( k \) arrives at the DCU.

A formal statement of the MSMMC problem is provided in the following definition:

**Definition 2.** Assume that over some period of time \( T \), all smart meters together generate \( n \) messages \( M = \{m_1, \cdots, m_n\} \) and send it through a \( b \)-ary data collection tree \( \Gamma \). At each DCU \( i \), \( i = 1 \cdot \cdots b^L - 1 \), in \( \Gamma \), \( n_i \) packets arrive from the lower level (either from other DCUs or directly from smart meters in the case of leaf DCU nodes). For a \( b \)-ary tree with \( L \) levels, \( n = \sum_{i=b^L-1}^{i=b^L-1} n_i \). For each DCU \( i \), the goal is to seek a \( r \)-partition \((r = 1 \leq n_i) P_{i1} \cup P_{i2} \cup \cdots \cup P_{ir} \) of the set \( n_i \) such that (i) each constructed packet meets the size constraint \( \sum_{k \in P_{ij}} s_{ik} + H_i \leq W_i, \ \forall j = 1 \cdot \cdots r \), and (ii) each packet \( P_{ij} \) meets the deadline constraint \( \max_{k \in P_{ij}} a_{ij} \leq \min_{k \in P_{ij}} d_{ij}, \ \forall j = 1 \cdot \cdots r \). A solution is optimal if it has minimal cost \( C \).
The MSMMC problem can also be formulated as a 0−1 integer linear program as:

\[
\text{minimize } C = \sum_{i=1}^{b_L-1} c_i V_i = c_i \left[ \sum_{j=1}^{n_i} y_{ij} \sum_{k=1}^{n_i} x_{ijk} \cdot s_{ijk} \right] + H_i, \tag{5.1}
\]

with \( j = 1 \cdots n_i, k = 1 \cdots n_i \)

subject to constraints

\[
\sum_{j=1}^{n} s_{ijk} x_{ijk} + H_i \leq W_i y_{ij}, \quad \forall k \in \{1 \cdots n_i\}, \quad i \in \{1 \cdots b_L - 1\}
\]

\[
\max a_{ijk} x_{ijk} \leq \min d_{ijk} x_{ijk}, \quad \forall k, j \in \{1 \cdots n_i\}, \quad i \in \{1 \cdots b_L - 1\}
\]

\[
\sum_{i=1}^{b_L-1} x_{ijk} = 1, \quad \forall j \in \{1 \cdots n_i\}, \quad k \in \{1 \cdots n_i\}
\]

\[
y_{ij} \in \{0, 1\}, \quad \forall j \in \{1 \cdots n_i\}, \quad i \in \{1 \cdots b_L - 1\}
\]

\[
x_{ijk} \in \{0, 1\}, \quad \forall k, j \in \{1 \cdots n_i\}, \quad i \in \{1 \cdots b_L - 1\}
\]

where \( y_{ij} = 1 \) if packet \( j \) is used at DCU \( i \), and \( x_{ijk} = 1 \) if message \( k \) is put into packet \( j \) at DCU \( i \).

### 5.4 Algorithms for the MLDA Problem

#### 5.4.1 Heuristic Approach

The MSMMC problem, as stated at the end of the previous section, falls into the class of online scheduling problems. A simpler version of the problem with only one DCU was proven to be NP-complete in the work of Karimi et al. [47]. Thus, in this work, heuristic algorithms are being developed for solving the MSMMC problem. The heuristic solution approach is to rely on EDF scheduling where a concatenated packet is created at the DCU based on the first expiring local deadline among all messages in its queue. Because the DCU could have both messages with and without deadlines, this EDF approach follows the following scheme for creating a packet: when it is time for the message with the earliest local deadline to be sent out, the DCU creates a new
packet with this message and fills any remaining space in the packet by a knapsack selection over all other messages in its queue. Evaluations [47] showed this heuristic outperforming five other EDF-based variations.

In the MSMMC problem considered, there are no best-effort packets at higher levels of the data collection tree; every packet created at the lowest level will have a deadline dictated by the EDF schedule. With multiple possible levels where concatenation can be done, schemes can differ based on the following criteria: (i) at what levels concatenation is done, and (ii) how local deadlines are computed at each level. In terms of the first criterion, schemes can choose to concatenate at only one level (lowest, highest, or somewhere in between), or at two or more levels. If concatenation is done only at one level, the surplus time remaining from a message’s deadline at the top level after accounting for processing and network delays at all levels can be applied to that one specific level for waiting to find a good fit that meets the required objective; packets created at all other levels will be simply forwarded by respective DCUs. If concatenation is done at multiple levels, then the time until the deadline expires at the top level for a message must be allocated in some manner to all levels after accounting for all processing and network delays expected.

Based on the above possibilities, in this work four different heuristics are proposed as listed in Table 5.1.

Table 5.1: PROPOSED MULTI-LEVEL AGGREGATION HEURISTICS

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDF-LLC</td>
<td>The concatenation process is executed only at the lowest level of the data collection tree.</td>
</tr>
<tr>
<td>EDF-TLC</td>
<td>The concatenation process is executed only at the root level of the data collection tree.</td>
</tr>
<tr>
<td>EDF-MLC</td>
<td>The concatenation process is executed at all levels, with overall deadline budget split among all levels equally.</td>
</tr>
<tr>
<td>EDF-WCF</td>
<td>The concatenation process is executed at all levels with overall deadline budget split based on a weighted function of costs associated at each level, as shown in equation (5.2).</td>
</tr>
</tbody>
</table>
For the heuristic EDF-WCF given that \( C_T = C_1 + C_2 + \cdots + C_L \), the waiting time \( (w_i) \) at each level \( i \) would be:

\[
w_i = \frac{C_i}{C_T} t
\]  

(5.2)

### 5.4.2 Reference Algorithms

The theoretical optimal method is the theoretically minimal number of packets that needs to go out of a DCU at each level for a given number of datagrams received from the lower-level data generators (smart meters or DCUs) over a period of time. This value is not constrained by arrival times or deadlines of messages; it is computed over all generated messages and maximum packet size MTU that includes a header size \( H \). For DCU number \( j \) at level \( i \), this value can be mathematically calculated by:

\[
Min(Packets_{ij}) = \sum_{i=1}^{h} \left\lceil \frac{\sum_{j=1}^{n_{c}(i)} S_j}{MTU_i - H_i} \right\rceil
\]

where \( n \) is the total number of arrived datagrams in a specific DCU during a time interval, \( S_d \) is the size of a datagram \( i \), and \( MTU_i \) and \( H_i \) are the parameters that vary depending upon the backhaul technology used at level \( i \). Although this solution is not feasible in practice, it provides a theoretical reference for the performance evaluation of MLDA algorithm, not limited EDF-based heuristics.

To obtain the minimum possible total number of packets at the highest level of data concentration

### 5.5 Evaluation

#### 5.5.1 Theory

To find the overall data reduction ratio \( \rho_{Total} \) through all levels of communication hierarchy the following equation is used:

\[
\rho_{Total} = \frac{\lambda_0 - \lambda_{h}}{\lambda_0}
\]

(5.3)
where $\lambda_0$ is the volume of data in bytes (including payload and control overhead) that would have been sent over the backhaul in a time period $T$ when message concatenation is not employed, and $\lambda_h$ is the volume of data sent (again including payload and control overhead) over the backhaul after last-level ($h$) packet concatenation.

5.5.2 Simulation Environment

The discrete-event simulator developed using MATLAB for evaluation of the proposed heuristic algorithms in the work of Karimi et al.[47] was extended to support a multiple-level architecture of message concatenation. This network topology consisted of a group of smart meters generating messages as Poisson arrival process at the leave level and sending messages to the primary level DCUs to be routed to the control center. Due to the assumption of each meter generation as a Poisson process, one can sum individual average message generation rates to get an average arrival rate at the DCU of $\lambda$, which is used as a parameter. This work only considered the $\lambda$ value of 1 that corresponds to 900 smart meters sending 1 message on average every 15 minutes. For the concatenation scheme also, only the best-performing heuristic, EDF-KN, proposed in the preliminary work of Karimi et al.[47] was chosen. The service capacity of the DCU was considered to be infinite. The height of the tree was assumed to be 2 (2-ary tree), and thus there were total of three DCUs at the simulation scenario: one at the root level and the other two at the primary level (level 2). The simulation time interval was kept at 10 seconds.

5.5.3 Distribution of Message Types

During a day, different types of messages may be exchanged between smart meters and the utility control center through the AMI. For these evaluations, all seven basic types of messages, listed previously in Table 4.1 and first reported by Luan et al. [62] were considered. Based on geographic location, power distribution infrastructure, and utility preferences, the transmission of

---

14Prior work [57] supports this assumption that smart meters message generation can be modeled as a Poisson process.
messages could come from different distributions of these basic message types, which all have an impact on the performance of the proposed heuristics. In the evaluations, different Beta distributions across these message types were used by varying the shape parameters, $\alpha > 0$ and $\beta > 0$.

Assuming that the arrival probability of different message types is $p_1, p_2, \ldots, p_n$ with $0 < p_i \leq 1$ for $i = 1 \cdots n$ and $\sum_{i=1}^{n} p_i = 1$, each can be computed as the result of the difference between cumulative probabilities of $n$ intervals as:

$$p_i = \text{betacdf}(i) - \text{betacdf}(i - 1)$$

where $\text{betacdf}$ is the cumulative density function of the beta distribution used with parameter, $\alpha$ and $\beta$.

For the experiments five different message type distribution were generated using the same shape parameters previously shown in Table 4.3 to test the performance of the proposed algorithms.

### 5.6 Simulation Results

On hundred runs of simulations were conducted with mean values along with 95% confidence intervals plotted in the results. Each scheme was evaluated in terms of the overall reduction in cost of data transmitted out into the uplink network by each DCU, as compared to the overall incoming data cost being only forwarded from smart meters to the upper level DCUs and Control Center, including all headers. Each packet header was assumed to be a fixed size of 50 bytes corresponding to the 40-60-byte range for TCP and IP headers. Figure 5.4 displays the output of the proposed algorithms and reference algorithm over five message-types distributions with 95% confidence intervals. Results are shown for packet arrival rates at the DCU of $\lambda = 1$. It can be seen that overall cost reduction varies 2-30%, depending on the message-type distribution, cost function order(CFO) at at each link, and specific algorithm used. From these results, three questions are answered: (1) How effective are the proposed heuristics? (2) What is the impact of message-type distribution? and (3) What is the impact of cost functions?
5.6.1 Effect of Proposed Heuristic Algorithms

The bar charts in Figure 5.4 show that the EDF-MLC algorithm has the best performance among all other heuristic algorithms and comes very close to the performance of the theoretical optimal across all CFOs and message-type distributions. This is because that EDF-MLC is using the original deadlines over all levels, provides more options to maximize the packet size before it is sent out. Since typically there are enough queued messages before a deadline is reached, the algorithm has a good collection of options to maximize the packet before sending it out.

Likewise the previous work [47] for a single layer of message concatenation where results been best when the arrival distribution was More smaller-sized, at Multiple-level data aggregation this distribution is also best suited for all of the heuristic algorithms. This is specially raises up due to the addition of cost factor at each level of tree structure.

5.6.2 Impact of Message-Type Distribution

The uniform distribution of all message types serves as the reference case to compare with other distributions. For the “more deadline” case where the majority of all messages have smaller sizes, the overall cost reduction is greater for all algorithms. The presence of more messages with smaller sizes than larger sizes assists packets to be sent out of the DCU with having more options and chances of the right combination to maximize packet size. On the other hand, larger packets present, algorithms are getting less options for the concatenation process before being forced to send out packets; Thus, when there are cost associated at each link, this causes more available messages being concatenated and therefore the overall data being transferred through the communication links is reduced which leads to greater cost reduction. The case for having more deadline restricted messages are similar to the more deadline message case in that it helps reduce packet overheads significantly through concatenation as header sizes are comparable to data sizes. Smaller messages are also easier to group into a packet. Conversely, the more larger messages case results in greater difficulty to group messages into a packet; also larger underlying size of messages already have reduced overhead making much improvement through concatenation difficult.
Figure 5.4: Overall cost reduction percentage using proposed MLDA heuristics over different cost order functions and message type distributions.
5.6.3 Impact of Cost Functions

Associating cost to the links at each level of b-ary tree topology of DCUs has an important impact on the data concatenation process performance. Given the results of multi-level data aggregation, it has been learned that depending upon the cost distribution (being in ascending or descending order from the root) over different levels of communication, the performance of MLDA could be drastically different. There are three different possible options for cost distribution: ascending order, descending order and equal cost distribution. Each of these cases is possible, depending on the definition of the cost by communication links design factors like technologies being used, traffic type, etc. To have a thorough and comprehensive analysis, in the simulations an example of all cases is considered. Figure 5.4 presented previously shows that it is always better to do the concatenation at the very first (bottom) level if not at all, no matter what cost function is being used. This is due to the fact that performing the concatenation at primary level DCUs is having the luxury of waiting for more time (the original deadline) to find the right combination for maximized-sized outgoing packets, whereas performing the concatenation at the root level is always of least benefit, due to lack of enough time for more processing and for considering options. At the same time, having the communication cost with descending order from root to leaves could downplays the above-mentioned fact and cause it to perform better compared to the case where the order would have been ascending. The weighted cost approach, EDF-WCF, however, is performing the concatenation at all levels, but still is not as good as multiple-level concatenation heuristic, EDF-MLC, because it is restricting the concatenation process at each level with a tighter deadline compared to the original deadline. The EDF-MLC always follows the original deadline.

5.7 Conclusion

In this work, the problem of multiple-level data message concatenation and aggregation in DCUs is investigated. First, the problem was defined based on requirements of smart meter applications such as AMI. Then it was mathematically formulated as an optimization problem whose solution fulfills requirements and includes appropriate concentration policies. The model is
designed as mixed-integer linear programming approach, which maximizes the size of the packet given the design objectives mentioned earlier. Therefore, the stochastic message arrivals will be considered through simulation tools such as Matlab or Arena, and the results will be shown based on numerical evaluations. The desirable outcome would be the outgoing packets with optimal sizes being sent out to uplink each DCU over all b-ary tree structures of DCUs, which is leads to the reduced cost of communications in terms of time, less overhead, and efficient load on the available capacity of the communication infrastructure. This will be compared with the case where no DCU is utilized in communications or DCU is utilized by a different mean and only forwards the messages to the next hop as a router. Also, it is possible in some cases to have only suboptimal policies.
CHAPTER 6
DISCUSSION, CONCLUSIONS, AND FUTURE WORK

6.1 Discussion

This section of the dissertation briefly discusses how the contributions of this work can be applied in practice for smart grids. This first part of the dissertation provided a comprehensive study of the requirements for a communications network at the distribution level and what candidate technologies exist to meet these requirements. These results would be very useful in the planning of a communications infrastructure to meet specific needs for a utility and actually knowing what performance to expect. Even if technologies of the future may be different, some requirements and possible deployment architectures would be similar, and will serve as a useful guide for further planning studies.

Given the nature of the physical topology of the power distribution system, there have been many proposals to reuse the power lines and associated infrastructure to build communication networks. Long, linear, wireless communication networks would become more likely under these scenarios and warrant more extensive studies than what has been done in the literature. The work in the second part of the dissertation on studying the capacity of linear multi-hop wireless networks will be very useful in understanding the capabilities and limitations of such a network using off-the-shelf WiFi radios. The analytical study presented can help estimate parameter values required for practical deployments including the number and type of linear chains required.

The third part of this dissertation proposed data concentration algorithms that can be directly employed in DCUs to solve network congestion and/or minimize data-gathering costs. These algorithms were designed based on current standards for meter data collection and, hence, should be directly applicable to deployments. Any such deployments using these algorithms should consider and test for the impact of variability in network latencies and DCU capabilities to obtain a better sense of expected benefits.
6.2 Conclusions

This dissertation considered the design and analysis of communication networks and associated mechanisms for the distribution system of smart grids in the future. It was divided into three major parts, each making a major contribution to the body of research on communication networks for smart grids. These parts have tried to answer the following high-level questions:

1. What is the minimum data capacity required to achieve a guaranteed level of performance in a power system?

2. What is the maximum data-carrying capacity given a fixed real-world power system infrastructure?

3. How can the available network capacity be utilized efficiently to handle the Big Data injected into the network through applications such as AMI?

Chapter 2 studied possible communications technologies and compared their capabilities with an emphasis on why wireless communication technologies might be best suited for the communication infrastructure. A wireless mesh network architecture was proposed and shown to satisfy the requirements of the power distribution system through protocol selection, tuning, and extensive simulations.

Chapter 3 considered the capacity planning aspect of backhaul networks likely to be used in power distribution systems. This was specifically studied through the advanced metering infrastructure application, which requires bidirectional communication between electric meters of customers and the utility control center. A linear chain multi-hop wireless communication architecture was proposed, and its ability to meet the application requirements of the communication backhaul were analyzed theoretically and evaluated through simulations.

Chapters 4 and 5 studied the issue of how to handle and transfer large volumes of data collected by electric utilities through resource-constrained communication networks. Heuristics were proposed to solve the problem of how to concatenate multiple small smart metering messages arriving at data concentrator units (DCUs) in order to reduce protocol overhead and thus network
utilization. These heuristics were shown to be able to reduce overall data volume by 10-25% for each DCU. This reduction was achieved just by a reduction in protocol overhead with no compression of the original data sent by smart meters, leaving much room to develop additional data concentration mechanisms at DCUs. Multi-level DCU scenarios were subsequently considered with additional heuristics proposed to take advantage of a network of DCUs over data-gathering trees.

6.3 Future Work

There is huge potential for additional research in area of communications network design for smart grids. Specifically for power distribution systems, the following research directions can be explored:

- **Evaluation of proposed heuristics on real deployments:** The proposed heuristics for data volume management, as discussed in Chapters 4 and 5, were evaluated through simulations. Although many practical considerations were taken into account, it would be interesting to study how variable network latencies, wireless communication characteristics, and implementation constraints may impact the results shown in this dissertation.

- **Data compression and aggregation:** In this dissertation, message concatenation was the only approach used for data concentration. There is great potential in reducing the volume of data by using data compression and data aggregation techniques.

- **Smart meter tamper detection and prevention:** With the transition to digital technologies and communication for collecting metering data, electronic tampering is likely to be of concern, in addition to the traditional concern of physical tampering. It is conceivable that malicious parties can modify/suppress two-way communications. It is also possible that some customers may modify their own smart meter’s readings to benefit from some of the proposed scheduling policies in this dissertation. Thus, the security and privacy aspects related to this dissertation would need to be explored in depth.
REFERENCES


[23] Cisco. Smart grid-transforming the electricity system to meet future demand and reduce greenhouse gas emissions, 2013.


[34] Evanczuk, Stephen. Data Concentrators Combine AFEs, MCUs, and Radios to Key Smart Grid Efficiency, March 2013.


APPENDICES
APPENDIX A
DISSERTATION-RELATED PUBLICATIONS


APPENDIX B

OTHER PUBLICATIONS

