5-2008

Environmental Observing Systems for Assessing Impacts of Land Use Change

Samuel Esswein
Clemson University, sesswei@clemson.edu

Environmental Observing System for Assessing Impacts of Land Use Change

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Forest Resources

by
Samuel Esswein
May 2008

Accepted by:
Dr. Christopher J. Post, Committee Chair
Dr. Sebastian Goasguen
Dr. Elena Mikhailova
Dr. Dan R. Hitchcock

ABSTRACT

The purpose of this research is to evaluate the application of technology solutions for enabling environmental research. This project develops a complete environmental observing system in support of an investigation of the hydrological and ecological impacts of land use change on the coast of South Carolina. The land use change study is an ongoing multi-disciplinary effort involving the collection of a large number of monitored and sampled parameters at Bannockburn Plantation, which is located near the City of Georgetown in South Carolina. Long term monitoring will support a hydrological and ecological assessment of the study site before, during, and after proposed residential and commercial development. There are three primary objectives of this work. The first is the comprehensive instrumentation of a portion of the Bannockburn study site for hydrological parameters. The second is the implementation of a wireless sensor network to support the remote acquisition of monitored data. The third objective is the creation of a reliable and robust software solution for transmitting and distributing real time observation data. Collectively, these three objectives provide a complete environmental observing system. Results of this work will benefit land use change research by providing access to real time observation data and enabling the integration of powerful analysis tools. This research will provide a basis for the development of future environmental observation systems.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Christopher Post for his guidance and enthusiasm for this research project. I would like to thank my committee members, Dr. Sebastian Goasguen, Dr. Elena Mikhailova and Dr. Dan Hitchcock for their support. Thanks to Dr. Tom Williams for sharing his vast hydrological knowledge and Josh Lawrence for his circuit design expertise.

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

PREFACE

This research explores the application of emerging technologies in support of a multi-disciplinary study investigating the impacts land use change along the coast of South Carolina. Research is organized into three general topics. The first topic, described in Chapter 2, explains the land use change study, its goals and the contribution of this research towards those goals. The second topic, described in Chapter 3, focuses on the design and implementation of a wireless sensor network developed to support remote acquisition of data for the study. The third and final component, described in Chapter 4, presents a software solution to manage and distribute observation data generated by the study. Taken collectively, these three topics comprise a complete environmental observing system. Results of this work may serve to guide for future implementations of environmental observing systems.

The land use change study incorporates a broad range of ecological and hydrologic research efforts. Research towards this thesis is focused primarily on the ground and surface hydrology components. The nature of the planned larger study requires long term data collection over a period of 3 to 15 years. This time span is necessary to capture changes invoked by proposed residential and commercial development on the site. The study is currently early in its design and implementation phase, so data are not in sufficient quantity or quality for useful analysis. As a result, Chapter 2 focuses on the experimental design, implementation and soil data processing.

The remaining chapters introduce methods for collecting, managing and distributing observation data. While these methods are presented in the context of the land use change study, they have applicability to a much wider range of environmental monitoring efforts. A key motivation for this work is the need for common infrastructure approaches in environmental observing systems. As a result, scalability and flexibility were key requirements throughout the design process. Chapter 3 presents a flexible, low-cost wireless sensor network capable of supporting heterogeneous instrumentation needs. Chapter 4 introduces an innovative approach to managing large volumes of streaming real-time data.

CHAPTER TWO

HYDROLOGIC MONITORING FOR LONG TERM ASSESSMENT OF LAND USE IMPACTS FROM COASTAL DEVELOPMENT

Introduction

This research is a component of a large interdisciplinary effort to evaluate the impacts of land use change on hydrology and ecology in coastal areas. This discussion focuses on the surface and groundwater hydrology component of the project. Research occurs on the Bannockburn Plantation (33.391º, -79.176º), a currently undeveloped 3,500 acre (1,416 hectare) tract located near Georgetown, SC (Figure 2.1). This site is also referred to as Arcadia East because it shares landowners with the Arcadia Plantation to the west. Currently, the site is characterized by upland pine and forested wetlands. It is privately managed for recreational uses (e.g. hunting and fishing). Current plans show that, over the next three to fifteen years, the property is slated for development including a hotel with a golf course, a multi-family housing village, a retirement village, a single-family housing village, a corporate park, and a shopping area. A notable characteristic of the site is its proximity to the North Inlet/Winyah Bay National Estuarine Research Reserve (NERR). As a NERR site, North Inlet/Winyah Bay has been federally designated for long-term stewardship and research by the National Oceanographic and Atmospheric Administration (NOAA). North Inlet and Winyah Bay are also considered to be “Outstanding Resource Waters” (ORW) as designated by the State of South Carolina. The study site contains a portion of the Debidue watershed which drains into the North Inlet. Debidue Creek was listed in 2006 as a 303(d) impaired watershed for fecal coliform concentrations [1], although standards for ORWs are typically much lower than those of typical sample locations. Results from this study may impact future decisions for water management practices along coastal streams and in coastal watersheds.

The shallow water table and low topographic relief of the Bannockburn site are typical of the coastal areas of South Carolina. These characteristics present a unique challenge for hydrological study and water management. Surface, sub-surface and ground water movement is tightly coupled, with close interaction between the component systems. A comprehensive monitoring program is used to gain a better understanding of the influences involved in the hydrologic systems of coastal South Carolina. The long term sampling and monitoring effort will provide an assessment in the context of pre-, during and post development conditions. Measurements include groundwater level, surface water discharge, rainfall, rainfall canopy through fall and weather conditions. Based on instrumentation, factors like evapotranspiration, rainfall-runoff and infiltration rates can then be determined and incorporated into computer based models. Future work may lead to improved storm water management practices. In particular, study results may determine the suitability and improve upon existing watershed evaluation tools based on SCS Curve Number based modeling practices (e.g. TR-55 and TR-20 [2]).

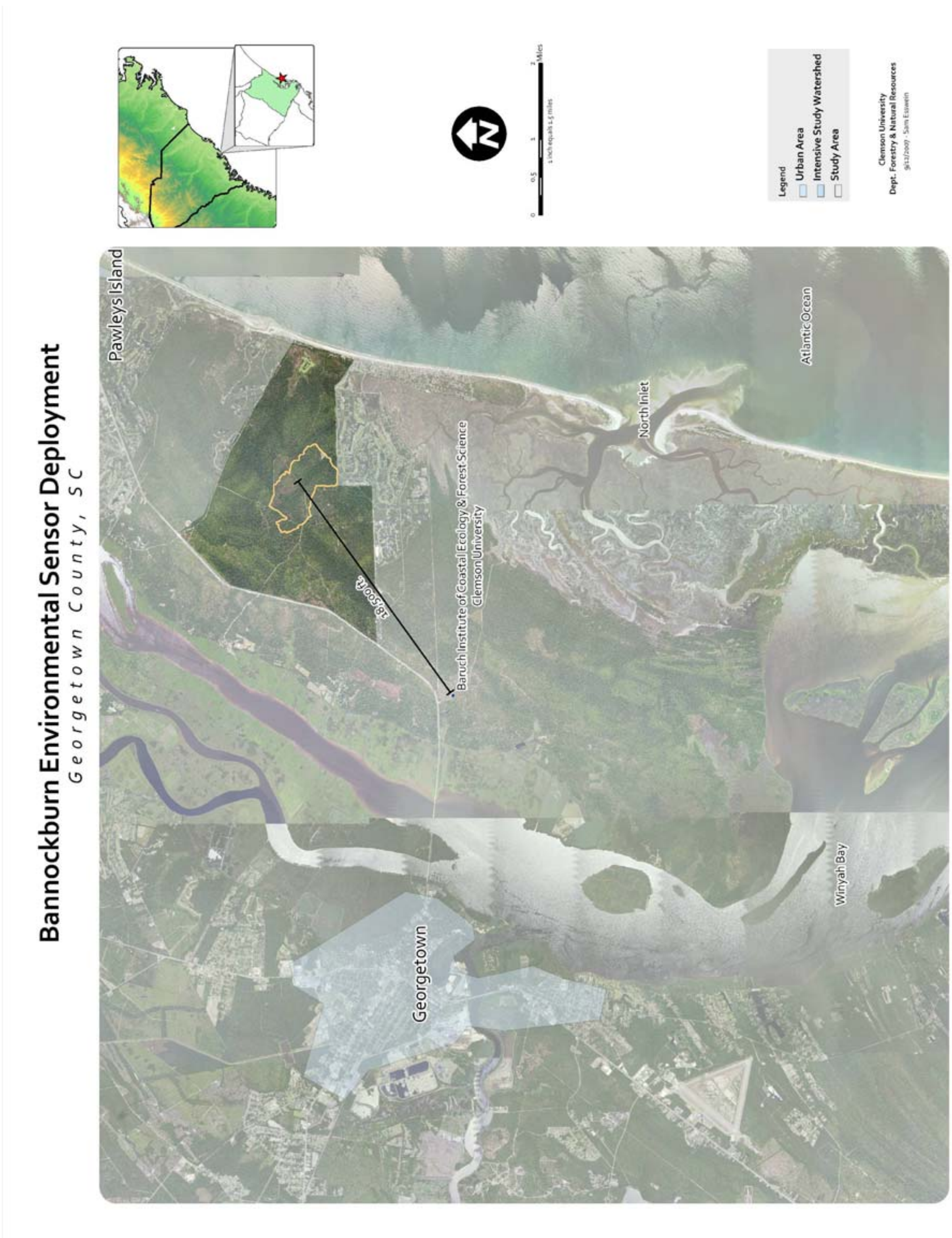


Figure 2.1 - Study Location

This study is currently implemented as a two phase effort. This discussion focuses on the first phase of implementation which is focused on a subwatershed area of Debidue Creek (outlined with orange in Figure 2.1). A detailed description of the study design and sensing instrumentation is included in the methodology section. This includes piezometer wells, weather, rainfall measurement and acoustic Doppler/pressure transducer based surface water discharge measurement.

Methodology

Phase One Study (Intensive Watershed)

The first phase of Bannockburn research focuses on a subwatershed (Intensive Study Watershed) of Debidue creek, as shown in Figure 2.2. This study area encompasses 337 acres (136 hectares) with elevations ranging from 6 – 24 ft (1.8 – 7.3 m). Weather stations installations are planned in the upper and lower sections of the intensive study area. Currently, only the lower weather station has been installed. Two weather stations offer both redundancy and the capability to evaluate microclimate variations within the study area. Weather station locations were chosen based on distance from trees and vegetation to ensure accurate precipitation and wind measurements. Surface water monitoring takes place at the upstream and downstream boundaries of the Debidue subwatershed. The monitoring and sampling locations chosen pass under a roadway through two corrugated culverts. The culverts offer a controlled channel

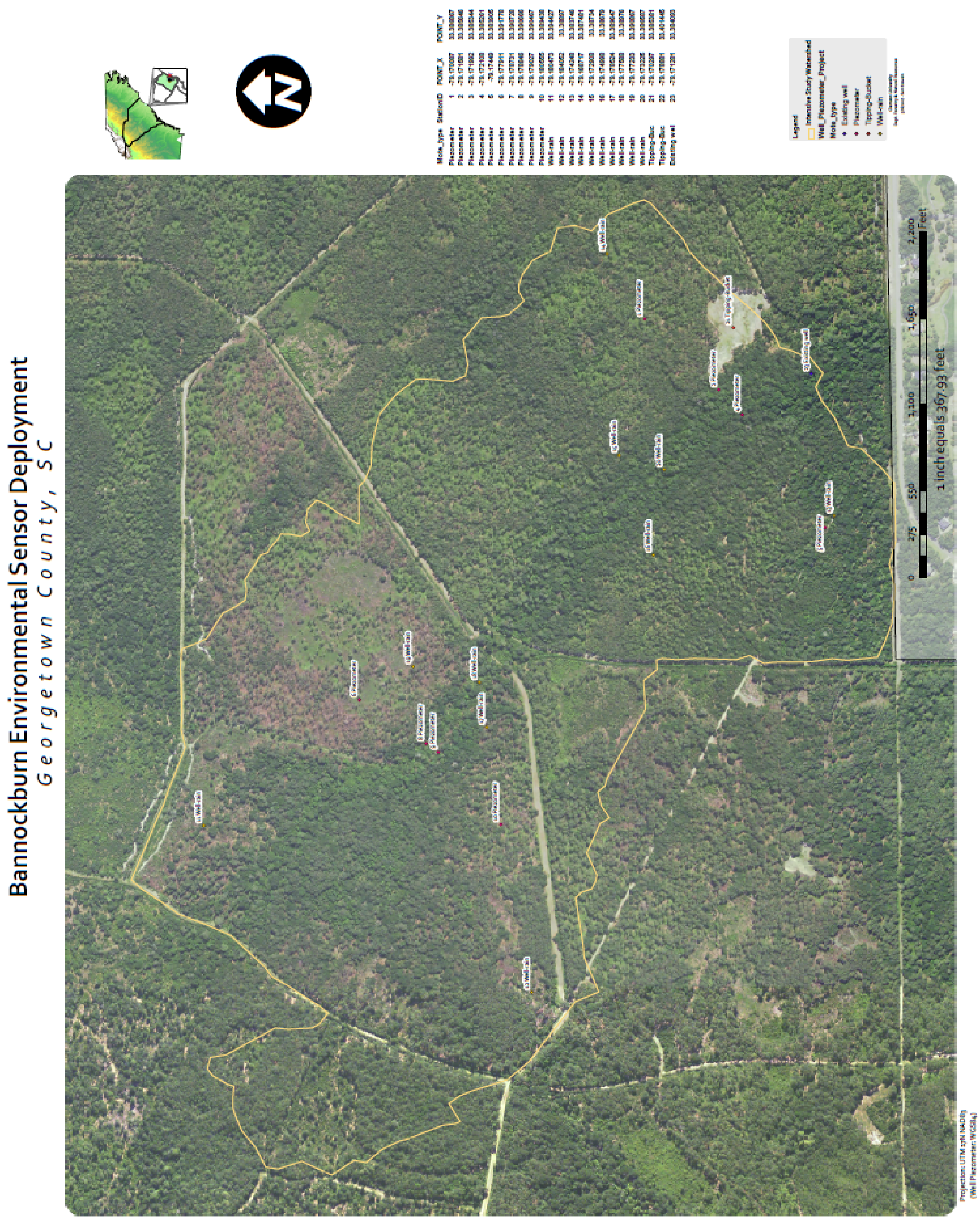


Figure 2.2 - Intensive Watershed Study

for discharge calculation. The multi-level piezometers support determination of ground-water input to stream flow. These piezometers are located along two transects located in the upper and lower portions of the subwatershed. Each transect has five monitoring locations positioned at elevations representing the watershed divide, the transition from pine to hardwood and the bottom of the stream channel. In addition to ground water level, the multi-level piezometers provide measurement of piezometric potential at various depths. A second type of piezometer station incorporates sensors to assess rainfall/canopy thru-fall, ground water level and soil moisture at the top of the B soil horizon. Locations for these sites were chosen to provide comprehensive coverage of the subwatershed area. A collection of soil samples is taken at each well location at 6 inch (15.24 cm) intervals. Each sample is processed for soil characteristics including pH, texture and composition.

Piezometer: Multi-Level

The multi-level piezometers utilize a three pressure transducer design that measures pressure at 15, 10 and 5 feet (4.6, 3.0, 1.5 m) below the ground surface. Pressure readings are taken every 30 seconds. Three differential pressure transducers provide the capability to measure piezometric potential at each depth. The piezometric potential will help to characterize the movement of ground water and determine its contributions to surface water discharge. An example of a multi-level piezometer is shown in Figure 2.3. Construction of the piezometer involves drilling a 4 inch (10.2 cm) boring to a 15 foot (4.6 m) depth.

Each transducer is located in a sealed and weighted 1 inch (2.5 cm) PVC pipe. The Measurement Specialties Level Master series of transducers are used for each depth reading [3]. The Level Master transducers use a supply voltage of 5V. The 5 foot and 10 foot measurements are carried out using a transducer designed to read from 0 – 5 psi. The 15 foot measurement is performed with a transducer designed to read from 0 – 15 psi. Using a 10-bit Analog to Digital Convertor (ADC) measurements of 0.011 of a foot (3.4 mm) are possible with the 5psi transducers. The 15 psi transducers have a resolution of 0.033 of a foot (10.1 mm) with a 10-bit ADC. Each transducer was tested under controlled conditions to generate a calibration curve for voltage to level conversion.

Figure 2.3 - Multi-level Piezometer

Transducers are vented to the surface using tubing to ensure the measurements are always taken against atmospheric pressure, even in the event of leakage or water accumulation inside the PVC tubing. Transducer cabling is run the length of the PVC and exits the top of the well through a 4 inch PVC well cap. The well cap is connected to the radio unit using a waterproof ½ inch conduit. In addition to the vent tubing and electrical connections, each transducer depth includes tubing for pulling grab samples for periodic conductivity measurements.

Piezometer: Canopy Interception and Evapotranspiration

Evapotranspiration piezometers consist of three types of measurement devices; a piezometer, soil moisture probe and a tipping bucket rain gauge. Pressure and soil moisture readings are sampled at 30 second intervals. The tipping bucket utilizes a revolving 24 hour precipitation total that is measured every 2 minutes. Precipitation measurements provide an indication of rainfall interception by the leaf canopy. The piezometer sensor is installed at a depth of 10 feet (3.0 m). A Measurement Specialties Level Master 5psi transducer is used for pressure measurement. The piezometer employs a 3 inch (7.6 cm) screened casing to minimize the movement of silt into the well interior. The transducer is sealed in a 1 inch PVC housing that runs the length of the well. Wiring and vent tubing pass through this housing. The top of the well is capped with 4 inch PVC and the cabling is run to the radio unit using ½ inch waterproof conduit. The soil moisture probe is an ECH_2O EC-5 sensor developed by Decagon Devices [4]. This probe is powered with a supply voltage of 3.3V. Soil moisture sensors are factory calibrated using soil samples collected form the research site. This process yields accuracy of 1-2% for volumetric soil moisture. The tipping bucket rain gauge is manufactured by Rainwise, Inc [5]. To lessen the likelihood of fouling from debris a metal grate was added to the design in order to catch debris before it enters the tipping bucket sensor. The tipping bucket is installed above the radio unit and is typically within 3 – 5 feet (0.9 – 1.8 m) of the piezometer and soil moisture probe.

Weather Station

Weather data is collected using a Rainwise MK-III RTI-LR sensor assembly, shown in Figure 2.4 [5]. The Rainwise unit measures wind speed, wind direction, temperature, relative humidity, barometric pressure and rainfall. Power for the unit is supplied through an attached solar panel. Weather readings are transmitted every 2 seconds.

Figure 2.4: Rainwise Weather Station

Surface Water Quantity and Quality

Surface water quantity parameters are currently being collected using a Isco 2150 area velocity sensor. Water quality parameters are collected using an In Situ ® Troll 9500 Professional XP sonde. Due to incompatibilities with the remote data acquisition system, this configuration will be replaced with an Isco 6712 automated sampler (Figure 2.5) with a Model 750 Area Velocity Flow Module [6]. The In Situ sonde supports direct connection to the Isco 6712. The Model 750 utilizes two sensors for calculation of flow rate. Average stream velocity is determined using a 500 kHz Doppler sensor. A second pressure transducer sensor measures water depth. Flow is calculated by multiplying the cross-sectional area of the flow stream in the culvert barrel by the average velocity.

Source: Teledyne/Isco

Figure 2.5 - Isco 6712 Automated Sampler

This approach allows greater flexibility in the event the sampling locations (culverts) become submerged or flow conditions are reversed. During the initial testing phase, level and velocity measurements have been taken every 5 or 15 minutes.

Discussion & Future Work

This research represents a system design and testing phase for an ongoing project focused on the long term monitoring of the hydrological and ecological impacts of land use change. The first phase of this research offered an opportunity to evaluate the technology and installation procedures for long-term study. In the next phase of implementation, the Middleton Creek watershed of Bannockburn Plantation will be instrumented using the methods developed and refined from this phase. The second phase will increase sensor density and incorporate a greater variety of measurement parameters. Future work will include an evaluation of models and analysis observation data.

References

[1] SCDHEC, "State of south carolina integrated report for 2006, path I: Listing of impaired waters." SCDHEC Bureau of Water, Columbia, SC, Tech. Rep. 06_303d, 2006. Available: http://www.scdhec.net/environment/water/docs/06_303d.pdf

[2] J. Werner, D. E. Woodward, Q. D. Quan, R. D. Nielsen and R. Kluth, "Estimation of direct runoff from storm rainfall," NRCS, Washington, DC, Tech. Rep. Part 630 Hydrology, July 2004,

[3] Schaevitz, "Level master: Low pressure isolated pressure sensors," Measurement Specialties, Hampton, VA, 2003.

[4] Anonymous (2008). ECH20 Probe Models EC-5. [Online]. Available: http://www.decagon.com/Ech2o/sensors1

[5] Anonymous (2008). RainWise, Inc - Industrial Precipitation Sensors. [Online]. Available: http://www.rainwise.com/products/

[6] Inc Teledyne Isco. (2008). Isco 6712 Full-Size Portable Sampler. [Online]. Available: http://www.isco.com/WebProductFiles/Product_Literature/201/Portable_Samplers/FullSize_Portable_Samplers/6712_Fullsize_Portable_Sampler.pdf

CHAPTER THREE

WIRELESS SENSOR NETWORK FOR ENVIRONMENTAL MONITORING

Abstract

Wireless sensor networks can benefit environmental monitoring efforts by supporting lower cost sampling at higher spatial and temporal resolutions over longer time periods. Many challenges must be addressed when implementing wireless sensor networks for environmental monitoring. Current technologies offer a wide range of possibilities for systems and networking configurations. This wireless sensor network implementation provides an evaluation of existing technologies and offers an innovative approach to remote data acquisition that addresses data transmission over large geographic areas.

Introduction

Wireless sensor networks provide enormous opportunities for environmental monitoring applications. Wireless sensor networks (WSN) support monitoring at high spatial and temporal resolutions over long periods of time. This in turn will support greater access to environmental data at lower costs. In certain cases, WSN’s may enable monitoring where it was previously impossible due to site inaccessibility or impacts from human presence. The benefits of WSN’s are not without challenges and cost. This discussion provides a description of an innovative approach that addresses the unique needs of an environmental monitoring application. This includes a methodology for physical implementation, sensing node and radio characteristics. An evaluation of the functionality of this sensor network implementation is provided, as well as a description of future plans for applications in environmental sensing.

Background

Wireless sensor networks are spatially distributed sensing devices. Devices are autonomous in the sense they are able operate independent of each other. They may rely on neighboring devices for communication and may intelligently react to events occurring elsewhere in the network. Devices are typically cost, size and power constrained depending on the nature of their implementation. These constraints typically influence and limit the amount of intelligence available on a given sensor node. Advancements in embedded computing support greater processing capabilities while lowering cost and power requirements. Despite these advancements there are still a significant number of challenges involved when choosing wireless sensor network technologies.

The design and implementation of a wireless sensor network for environmental monitoring requires balancing the characteristics of the monitoring effort with available technology and resources. As noted by Callaway [1] wireless sensor networks are faced with a variety of challenges not found with traditional communication infrastructure. Unit cost, power consumption, radio characteristics, processing power, memory and sensor characteristics must all be taken in account when designing a network implementation. Design decisions were governed by the unique monitoring requirements, an analysis of available technology and prior experiences with existing technologies.

Wireless sensor networks are a relatively new area of research and much of the available research and technology has been limited to academic or laboratory settings. Existing technology for sensor nodes and wireless communications are limited in terms of cost, power efficiency and reliability. This is particularly true when deployed in harsh and unpredictable conditions common to environmental monitoring applications. A notable study involved a habitat monitoring study on Great Duck Island [2]. This study was a landmark effort in wireless sensor networks for several reasons. First, it utilized intelligent sensing nodes based on the Mica platform of motes developed at UC Berkeley [3]. Second, the study involves power constrained nodes deployed over long periods of time. The study comprises an area of 237 acres (95.9 hectares) and has limited access to grid power.

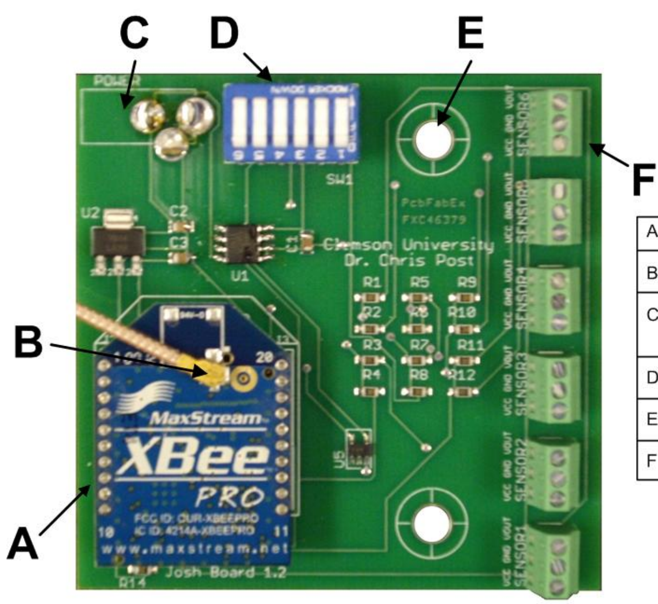
This wireless sensor network supports a research effort which aims to assess the impacts of land use change occurring along the coast of South Carolina. Rapid commercial and residential development impacts the hydrologic and ecologic systems of South Carolina’s traditionally forested coastal areas. The sensor network will monitor the condition of a 3,500 acre (1,416.4 hectares) tract prior to, during and after high-end residential and commercial development takes place. In-situ monitoring occurs through a variety of existing and purpose built instrumentation. The heterogeneity of instrumentation results in a wide range of interface and sampling requirements. Some of this instrumentation has been traditionally deployed using conventional data logging capabilities. Supporting remote data acquisition offers two significant advantages for this monitoring effort. First, remote data acquisition supports near real-time collection of observation data for support of richer analysis and response applications. Second, it removes the need for frequent site visits by a technician for data collection. This in turn allows us to monitor at spatial resolutions that may otherwise be limited by cost and personnel availability. A significant constraint with this sensor network design was the transmission distance requirements and the dense vegetation present throughout the sampling area. This presented a number of challenges, particularly in light of our budget constraints and quantity of sensing nodes.

Methodology

Analog to Digital Radio Unit

Maxstream/Digi [4] radio modules were chosen for radio communications. These modules include support for 10-bit analog to digital conversion (ADC). A custom board (Figure 3.1) was developed to support the ADC capability, provide ADC signal conditioning and to supply power for radio modules and sensors. This board supports six single-ended analog sensor connections (for full specifications see Table 3.1). Sensor connections must support three conductor connections in the form of input voltage (Vcc), ground (gnd), and output voltage

(Vout). The current board implementation supports input voltages of 5 volts. Output voltage supports scaling the corresponding digital output across either a 2.5 volt or 5 volt range. This feature is designed to allow the ADC to support a wider range of possible output voltages from attached sensors. Some sensors output voltages from 0 – 2.5 volts while others may require 0 – 5 volts. Output voltage adjustment allows a 0 – 2.5 volt sensor to utilize a larger portion of the 10-bit ADC range. This setting is available through a DIP switch as shown in Figure 3.1. When the DIP switch is set to off, the voltage range is from 0 – 2.5V.



A	Radio Module
B	Antenna: U.FL Coaxial RF
C	Power: Barrel Connector Male 2.1mm (Located on Board Underside)
D	Voltage Divider: DIP Switch
E	Mounting Hole
F	ADC: 3 Conductor Screw Terminal

Figure 3.1 - Analog to Digital Convertor Radio Unit

Table 3.1 - Analog to Digital Convertor Radio Unit Specifications

Specifications	
Input Power	5 – 14 V 2.1mm Barrel (Male)
Sensor	6 Sensor (VCC/GND/VOUT) 20-28 AWG
Sensor Input	5V
Sensor Output Range	2.5V or 5V
Physical	Width: 68mm (2.626in) Height: 68mm (2.626in) Depth: 25mm (.984in)

Physical/Implementation

Installation of remote acquisition nodes requires a weatherproof enclosure to contain the Analog to Digital Radio Unit and a battery. This enclosure is located above the ground and should be able to withstand the elements including rain, freezing temperatures, insects, animals and extended exposure to sunlight. The National Electrical Manufacturers Association (NEMA) has created standards for selection of enclosures for electrical equipment based on environmental conditions [4]. The enclosure shown in Figure 3.2 meets the NEMA 4x rating and has physical dimensions of 14” x 12” x 7” (355.6 x 304.8 x 177.8 mm). The dimensions are adequate to contain the analog to digital radio unit and the 5Ah battery detailed in the power section of this document. Liquid tight conduit connectors are used to make connections into and out of the enclosure boxes. Typical installation involves an antenna connection exiting the top of the enclosure and one or more holes in the bottom of the enclosure for sensor connections. All sensor connections encase the wiring with armored conduit to prevent damage from the elements, insects and animals.

Physical installation of enclosures involves a cementing an 8 foot (2.4 m) section of 1 5/8” (41.3mm) galvanized steel pipe 2 feet (.6 m) into the ground. The enclosure is mounted to a 1 x 2 foot (30.5 x 61.0 cm) section of 3/8” (9.5 mm) pressure treated plywood. This plywood is then attached to a 2 foot (61.0 cm) section of 2 inch (50.8 mm) PVC. The PVC section slides over the galvanized steel pipe and is secured using a lag bolt or drywall screw. The PVC section allows easy removal in the event a monitoring station must be removed, relocated or replaced. An antenna may be mounted to the plywood section using an antenna mounting plate.

Figure 3.2 - NEMA 4x Enclosure

Power

Providing adequate power is a key concern in any remote data acquisition project. Depending on the nature of the sensing node, we have adopted two approaches for power use. Typical low power nodes use a single 12V 5Ah rechargeable lead acid battery. For piezometer and soil-moisture sensing applications, this battery is sufficient to supply power for 4 to 6 months. It can be connected directly to the regulated power supply located on the analog to digital radio unit. In the event a 5Ah battery is not sufficient to supply power, we utilize a 40 watt photovoltaic panel for solar power collection and a 120Ah deep cycle marine battery. Because of higher wind loads on the panel, a pressure treated 4x4 post is used for mounting of the photovoltaic panel, as shown in Figure 3.3. The 120Ah battery is mounted inside of a battery enclosure which is located on the ground.

Figure 3.3 - Photovoltaic Panel

Communications

Radio communications among nodes in the wireless sensor network is provided through Maxstream/Digi based radio and cellular products. Our analog-to-digital radio units are based on the Maxstream/Digi 2.4GHz XBee-PRO (Series 1) modules [5]. The XBee-PRO is low-cost and has excellent power usage to transmission distance characteristics. Extensive configuration supports a variety of possible implementation designs. Notable features include support for user defined sleep periods and sampling frequencies. The sleep periods are critical for supporting low power operation necessary for long term deployment. For communication, the XBee-PRO supports a basic IEEE 802.15.4 [6] configuration as well as a ZigBee [7] configuration for peer to peer mesh networking. Current configurations do not support sleep in ZigBee modes, so we have limited our configuration to standard 802.15.4. In our ADC applications, we utilize a star network topology. Central nodes are constantly powered and receive observations from peripheral nodes. Each peripheral node has an address which uniquely identifies the data it sends to processing applications.

The data arriving at central nodes is bridged onto a cellular link through a Digi ConnectWAN cellular modem. The ConnectWAN provides reliable operation for long-term remote connectivity needs. Available in multiple interface configurations, we use a RS-232 serial connection for bridging to the XBee-PRO radio units. In addition to serial connectivity, the ConnectWAN provides Ethernet connectivity for simplified configuration and remote computer connectivity.

Cellular carrier was chosen based on service availability at the site and cost. Cingular/AT&T Communications [8] provides low cost unlimited data telemetry plans designed for remote data acquisition needs.

To maximize the range of our radio units, we utilize high gain omnidirectional antennas at our central and peripheral nodes. The peripheral nodes use a 12dBi gain antenna. The central nodes utilize a 15dBi gain antenna. To further boost performance at central nodes, a 1 watt amplifier is employed to increase power for transmit and receives.

Discussion

The greatest challenge faced during the implementation of this sensor network was the distances and dense vegetation between radio units. Traditional mote technologies like the one identified in the Great Duck Island study were determined to be inappropriate for our research site due to the large monitored area (3,500 acres or 1, 416.4 hectares). Incorporating mote technologies over this site would have required very high quantities of relaying motes and substantially increased the overall complexity of the network. Several alternatives were evaluated for radio communications. The best cost to performance options were based on Maxstream/Digi based radio products. The initial design called for 2.4GHz transmission between sensing nodes and a 900MHz link for backhaul communications to the Clemson University's Baruch Institute of Coastal Ecology and Forest Science. The 2.4GHz network is based on the Maxstream/Digi XBee-PRO series of radio modules. These devices advertise transmission capabilities of up to one mile. The backhaul communications were based on Maxstream/Digi Xtend 900MHz radios. The 900MHz radios advertise a 40 mile (64.4 km) range with a clear line of sight. Unfortunately no clear line of sight was available over the 5.5 mile (8.9 km) distance from the Baruch Institute to the research site. Testing indicated that the 900MHz link would not work as a backhaul option. As a result, cellular communications was chosen for the backhaul link between the Clemson campus network and our research site.

In the design phase of this project, we tested a variety of antennas, cables and radio units in order to evaluate the performance we could expect in our field implementation. The Maxstream/Digi 2.4GHz XBee-PRO RS-232 modem includes a half-dipole “rubber duck” antenna. Testing indicated that upgrading this antenna significantly improved radio performance. The full results of our 2.4GHz antenna tests are shown in Table 3.2. Tests were carried out in dense deciduous forest in an area with low topographic relief in order to mimic the conditions encountered at the research site. Transmission distances for the 2.4GHz antennas are significantly impacted by vegetation cover. An improvement of 389 feet (118.6 m) is accomplished when the default “rubber duck” antenna is upgraded to a higher gain omnidirectional antenna. Transmission distances do not improve significantly with the application of directional antennas. As a result, this implementation utilized high gain omnidirectional antennas for cost and complexity reasons. In addition to antenna configurations, antenna cabling tests were carried out. Cabling decisions can impact transmit and receive performance. Performance is related to cable length and cable quality. Lengths of one foot (0.3 m) and ten foot (3.0 m) patch cables were tested in two different cable qualities. Cable qualities were based on cost and Maxstream/Digi recommendations. Small improvements were noted with shorter cabling, but no detectable distance was found between the two cable qualities.

Additional challenges for sensor network implementation included physical installation considerations. Installation and operation of sensing equipment was made difficult by the inaccessibility of many of the sampling sites. Many sites were located in close proximity to water channels and/or in dense vegetation. To minimize the challenges of installation, repair and removal we designed our physical mounting and enclosures to minimize the amount of field installation time. Overall, the physical mounting and enclosure design has demonstrated its effectiveness in withstanding weather, animals and insects.

Table 3.2 - 2.4GHz Antenna Test Results

Base	Remote	Condition	Distance (ft)/(m)
B	B	Field / Deciduous	1,928 / 587.7
D	C	Dense Deciduous	1,300 / 396.2
B	C	Dense Deciduous	1,281 / 390.4
D	B	Dense Deciduous	1,120 / 341.4
B	D	Dense Deciduous	1,119 / 341.1
B	B	Dense Deciduous	1097 / 334.4
B	A	Dense Deciduous	708 / 215.8
A	**B**	**C**	**D**
Omnidirectional 2.4GHz 2.2 dBi Maxstream/Digi included w/ XBee-PRO RS232 modem	Omnidirectional 2.4GHz 12 dBi Gain Hyperlink Technologies PN# HG2412U	Sector Directional 2.4GHz 18 dBi Gain Hyperlink Technologies PN# HG2418P	Yagi Directional 2.4GHz 14.5 dBi Gain Hyperlink Technologies PN# HG2415Y

Power considerations were an additional challenge faced in this implementation. The research site offered no access to grid power. Power sources were limited to solar and battery power. The 40 watt solar arrays implemented for the backhaul and amplified central node have proved to be well matched to the power requirements of the devices. Sensing nodes are based on low cost 5Ah lead acid batteries. Laboratory calculations for battery lifetime expectancies indicated we should achieve our goal of 6 month deployment times for the analog to digital radio units. Accurate and reliable environmental monitoring results are not available due to the limited deployment times currently available.

Future Work

Future iterations of the analog to digital radio units will improve radio performance and offer additional capabilities for sensor flexibility, sensor resolution and intelligent sensing. This is achieved through the incorporation of an ATmega128 microcontroller and external 1-wire analog to digital convertors (ADC). A microcontroller will support improved sleep modes, adaptive sampling and intelligent observation aggregation and summarization. The use of external ADC units will offer flexibility in terms of ADC resolutions, including resolutions of 16-bit and greater. By removing ADC pin requirements from the radio interface, we gain greater radio module flexibility and the ability to upgrade to Maxstream/Digi's Series 2 XBee-PRO chips which have improved power efficiency and radio performance. Future testing may reveal opportunities for mesh networking approaches that offer a more resilient network topology and simplify multi-hop routes.

Conclusion

This chapter provided an overview of the challenges faced with wireless sensor networks implementations for environmental monitoring applications. An explanation of wireless sensor network technologies is provided, including an evaluation of current offerings. A complete remote data acquisition approach is presented; including physical implementation, radio considerations including frequency, antenna and cabling decisions. Challenges include the spatial and vegetation density characteristics of the site, from both a radio and installation perspective. Also included is a proposed direction for future research based on sensing implementations utilizing devices with greater processing and sensing capabilities.

References

[1] E. H. Callaway, *Wireless Sensor Networks: Architecture and Protocols.* Boca Raton, FL: CRC Press LLC, 2004,

[2] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk and J. Anderson. Wireless sensor networks for habitat monitoring. Available: http://doi.acm.org/10.1145/570738.570751

[3] Hill, J., Culler, D., "A wireless embedded sensor architecture for system-level optimization," Tech. Rep. UC Berkeley Technical Report, 2002.

[4] Anonymous (2008). Maxstream Radio Modules by Digi International. [Online]. Available: http://www.maxstream.net

[5] NEMA, "*Enclosures for electrical equipment (1000 volts maximum)*," National Electrical Manufacturers Association, Rosslyn, Virginia, Tech. Rep. 250-2003, 2003.

[6] Anonymous (2008). XBee-PRO 802.15.4 OEM RF Modules. [Online]. Available: http://www.digi.com/products/wireless/point-multipoint/xbee-pro-series1-module.jsp

[7] IEEE, "Wireless medium access control (MAC) and physical layer (PHY) specifications for low rate personal area networks (WPAN)," IEEE Computer Society, New York, NY, Tech. Rep. Std. 802.15.4 - 2003, 2003.

[8] ZigBee Alliance, "ZigBee protocol specification," ZigBee Alliance, San Ramon, CA, Tech. Rep. 053474r17, 2008.

[9] Anonymous (2008). Wireless Service from AT&T. [Online]. Available:

http://www.wireless.att.com

CHAPTER FOUR

APPLICATION OF PUBLISH/SUBSCRIBE MESSAGING FOR MANAGEMENT OF STREAMING ENVIRONMENTAL DATA

Abstract

Data stream management is a crucial component of environmental observing systems. Advancing technology (e.g. sensor networks) is supporting the integration of sensing instruments into the fabric of distributed systems. No longer sensing in isolation, sensing devices are increasingly relied on for real-time generation of data. In this paper, we explore the application of the distributed publish/subscribe messaging infrastructure NaradaBrokering for transmission of observation data streams. We introduce a complete messaging solution that leverages this messaging substrate to support the management of observation data streams with an appropriate data model and associated internet applications. To demonstrate the applicability of our approach to environmental observing, we introduce a ground and surface water monitoring effort for an investigation into the impacts of land use change in coastal regions.

Introduction

The management of streaming observation data poses a significant challenge to environmental observing systems. The generation of environmental observation data is increasingly exposed as real-time data streams. These streams may originate from any number of technologies e.g. wireless sensor networks, remote sensing platforms, or mobile devices. Sophisticated streaming data management supports the linking of streams from these technologies to consuming entities. The possibilities for data consumers are endless e.g. real-time analyses, network health monitoring, and event-response tools. Challenges stem from the heterogeneity of producers and consumers, the quantity of streams and the sheer volume and diversity of possible environmental observation data. As the size and complexity of these systems grow, so does the need for a scalable, reliable and high performance solution for managing the data streams. Supporting the efficient transmission and distribution of observation data requires a sophisticated communications and observation data management approach.

Typical environmental observation systems monitor the condition of the environment through the collection and communication of specific parameterized data. They support a broad range of research efforts e.g. weather predication, land use and climate change. The implications of this research can have far reaching significance for policy decisions, disaster response and sustainable environmental practices. Advancements in environmental observing systems have largely been led by the need to bring historically isolated observation systems together. The net result is the ability to study interactions between different types of observations and support larger scale efforts. These collaborative efforts have offered unprecedented opportunity for geographically and organizationally distributed researchers to share resources and coordinate efforts. Due to their size and scope, large environmental observation systems are faced with hundreds or thousands of potential data streams. We illustrate the application of our solution with an ongoing environmental monitoring effort taking place on the coast of South Carolina. This monitoring effort investigates the impacts of land use change in rapidly developing coastal areas. We incorporate a variety of heterogeneous instrumentation into a wireless sensor network that supports near real-time data collection and transmission.

Streaming data management systems benefit from sophisticated message-based approaches to communication. There are many factors to consider when selecting an appropriate communications approach. Our solution leverages a distributed messaging infrastructure based on the brokered publish/subscribe approach to communication. This approach facilitates a high degree of decoupling between observation data stream producers and consumers. The ability to reduce coupling is well suited to the dynamic and unpredictable nature of highly distributed observing systems. We couple our publish/subscribe approach with a powerful approach to observation data management. We introduce a suite of applications for processing observation data onto and off of the messaging substrate. To leverage existing applications in the earth and atmospheric science domain, we apply existing conventions and standards whenever possible.

Background

Environmental observation systems monitor the state of our environment. They support a broad range of research efforts e.g. weather prediction, land use and climate change. The implications of this research can have far reaching implications for policy decisions, disaster response and sustainable environmental practices. Advancements in environmental observing systems have largely been led by the need to bring historically isolated observation systems together. This allows a better understanding of the complex interactions between earth systems. The Geosciences Network (GEON) [1], National Ecological Observatory Network (NEON) [2], and Water and Environmental Research Systems Network (WATERS) [3] are just a few examples of developing regional- and continental-scale environmental observing systems. These collaborative efforts offer unprecedented opportunity for geographically and organizationally distributed researchers to share resources and coordinate efforts. Due to their size and scope, large environmental observation systems may support data collection from hundreds to thousands of potential data sources. A management approach needs to be flexible to deal with a) variations in data representation, b) expected lifetimes and c) acquisition/access methods. Our approach tackles these three constraints.

In large environmental observation systems, management and storage of data can be accomplished through either a centralized or federated approach. In the latter case, administration and authority is autonomous at an observation provider. Every organization may choose to handle their data collection and distribution in a different way. Accommodating the diverse requirements of organizations and users is a key motivator of our approach. This approach has been applied in support of a variety of monitoring efforts taking place through Clemson University. For this discussion we highlight an ongoing land use change study taking place on the coast of South Carolina. This monitoring effort was chosen for inclusion in this discussion for two reasons. First, a wireless sensor network is utilized to support near real-time data collection and transmission. Second, the sensing fabric is comprised of a wide variety of instrumentation.

This study is funded through the Program for Integrated Study for Coastal Environmental Sustainability (PISCES) [4]. Like many coastal areas in the United States, the coastlines of South Carolina are facing a high rate of commercial and residential development. This development has a broad range of implications e.g. water quality, coastal flooding and ecological health. Monitoring takes place on 3,500 acres (1,416 hectares) located to the north of the North Inlet/Winyah Bay National Estuarine Research Reserve (NERR). As a NERR site, North Inlet/Winyah Bay has been federally designated for long-term stewardship and research by the National Oceanographic and Atmospheric Administration (NOAA). The study tract (Figure 4.1) is currently undeveloped and privately managed for recreational uses e.g. hunting and fishing. Over the next three to fifteen years it will be developed to include a hotel with a golf course, a multi-family housing village, a retirement village, a single-family housing village, a corporate park, and a shopping area. This study monitors the pre-, during- and post-development impacts on hydrological and ecological conditions. A wide range of parameters are collected using a variety of sensing instrumentation (Table 4.1). Research findings will support, among other things, the development of improved hydrological and ecological management practices for coastal areas.

Table 4.1: Monitoring Parameters, Instruments and Quantities

Parameter	Instrument	Phase	Quantity
Surface Discharge	Acoustic Doppler/Isco 6712 – 802.15.4	1	2
Piezometer	Custom ADC/2.4GHz – 802.15.4	1	8
		2	13
E-T Monitor	Custom ADC/2.4GHz – 802.15.4	1	10
		2	17
Weather Station	2.4GHz – 802.15.4	1	2
Soil Moisture	Custom ADC/2.4GHz – 802.15.4	2	19
Dendrometer	Custom ADC/2.4GHz – 802.15.4	2	12

Wireless sensor networks (WSN) provide powerful data collection and transmission capabilities. In particular, WSN’s support long-term collection of observation data at high spatial and temporal resolutions. This WSN uses data acquisition devices developed at Clemson University. These devices emphasize low-power consumption and communicate using Digi/Maxstream based 2.4GHz radio modules [5]. We interface with both data-logging devices and custom analog-to-digital convertor (ADC) based sensing equipment. For example, groundwater parameters are collected via an ADC attached to a piezometer.

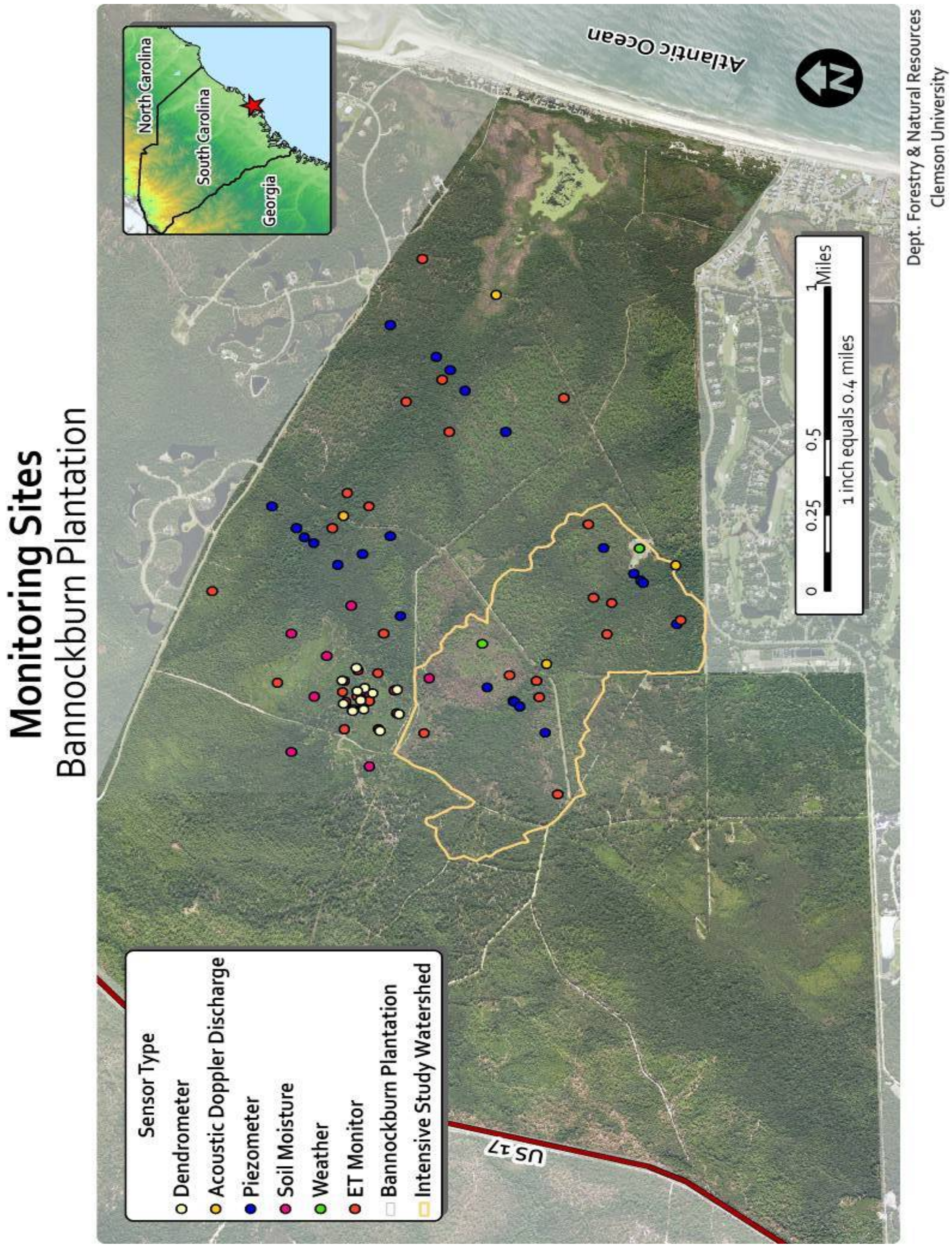


Figure 4.1: Monitoring Locations and Parameters

On the other end of the spectrum we have data-logging devices that require custom integration dependent on manufacturer specific application programming interfaces. For example, surface water discharge will be obtained using an Isco 6712 with a 700 series acoustic Doppler velocimeter/pressure module. This device relies on a high-level interface for management and data retrieval. Data retrieval involves “pulling” data through remote requests occurring at regular intervals. Our ADC-based devices “push” observation data at user-specified intervals (configurable at run-time).

Communications Architecture

Internet computing has created the opportunity to build distributed applications at unprecedented scales. The inherently dynamic and unpredictable nature of large distributed systems can hamper the development of sophisticated applications. Decoupling interactions between participating entities can help to reduce inter-dependencies and create more resilient applications. A decoupled or “loosely-coupled” interaction is an often cited benefit of service oriented architecture (SOA). In the following section we use three dimensions of decoupling defined by Eugster et al. [6] for characterization of communication interaction patterns. The first dimension, space, refers to the anonymity of interactions. Participating entities need not be aware of the other. The second dimension, time, specifies that sending and receiving parties need not be present on the system at the same time for the interaction to take place. Finally, synchronization decoupling implies that the program flow of participating parties is not impacted by the communication interaction.

The majority of communication interactions take place as a request/response. Typically, a client generates a request to a server and the server responds to the request. This approach is well-suited to client/server architectures where implicit response time constraints are beneficial. In most cases, program flow at the client is blocked until a response is received. In the event the response does not occur, the client must explicitly deal with the communications failure. Two notable request/response examples include; Hypertext Transfer Protocol (HTTP) and traditional Remote Procedure Call (RPC). Both of these examples are said to be “coupled” in time, space and synchronization.

Synchronization decoupling or asynchronous communication is frequently achieved through notification-based interaction. Notification is commonly referred to as publish/subscribe or as the observer pattern in object-oriented programming [7]. Asynchronous communications are considered a key component of grid technologies. The Globus Alliance has identified notification schemes as a component of their Open Grid Services Architecture (OGSA) [8]. Conventional forms of notification achieve synchronization decoupling but do not address coupling in space or time.

Achieving space and time decoupling typically requires the presence of a third party messaging intermediary or broker. This intermediary is commonly referred to as a message oriented middleware (MOM) and supports communication through the exchange of messages or events. In addition to routing messages, a MOM may include capabilities to support persistence of messages, message routing and data model translation among communicating entities. Our application focuses on a message oriented middleware that provides a brokered publish/subscribe communication pattern. This approach provides space, time and synchronization decoupling. This pattern is comprised of three entities: a publisher, a subscriber and a broker. Communication occurs through the exchange of messages between these entities. Messages are published to a broker under a particular topic or content descriptor. The broker (or collection of brokers) is responsible for routing the message to subscribers who have registered an interest in a given topic or descriptor. There are a variety of possible mechanisms for constraining or filtering messages. The most common is a string expression, but more advanced methods include regular expressions and XPath queries.

Data Models

With a multitude of data being collected by our sensors, there are clear advantages to adopting a “common ground” for representation of data and metadata. Data representation conventions support richer applications and decrease the complexity associated with supporting disparate data providers. A message oriented middleware can facilitate this by providing message translation capabilities. Data and metadata can pass between communicating entities despite differences in their respective data models. There are implicit assumptions about what data models are supported and not all models are compatible. For the purpose of this discussion, a data model describes a particular representation and way of accessing data. A data format describes the persistent form of a data model, either for network transmission or local storage. A data convention specifies the semantics of data, i.e. required data and metadata and its layout. Data model, format and convention choice is generally governed by its suitability for the data being represented and its application domain. Some models are geared specifically for point/station observations (typical of in-situ collection) while others are suited for gridded data (typical of remote-sensed). In the earth and atmospheric sciences there are a multitude of available data models and data formats. However, at least in terms of observation data, these models typically share enough commonalities to support meaningful translations. I.e. most widely used models support the functional equivalent of variables, attributes, structures and sequences. Ideally, a unified data model can support multiple models/formats without data loss.

The middleware adopts a native model and format for transmission of observation data and metadata. A producing entity can send data using any supported model. At some point, the middleware translates the model into the native model. One or more data consumers receive data using the data model of their choice. A native model supports the standardization of tools for observation processing. For example, in-situ station observation data is expected to have a spatial component. Therefore, a station observation processor should be able to understand the spatial location of all station observation data.

The capability to perform data translation opens up many opportunities for simplified storage, analysis and visualization. Coupled with multicast support, a single data source may support any number and variety of potential uses. However, this approach may not be appropriate in all environmental observing scenarios. There are drawbacks from a complexity standpoint. From a coupling perspective, one can argue that implicit assumptions about model compatibility lead to tighter coupling of entities. However, achieving decoupling in space will always require mutual understanding about data representation between entities. Furthermore, complexity is significantly alleviated through the adoption of a service oriented approach and the use of a message oriented middleware.

Methodology

NaradaBrokering

Our streaming observation data management system is based around the brokered publish/subscribe paradigm. There are a number of message-oriented middleware solutions that support this communication pattern. Our choice is governed by additional needs for scalability, performance and reliability. NaradaBrokering was chosen for its suitability for this type of grid oriented application. NaradaBrokering is an open source, general purpose messaging solution developed in Java by the Community Grids Labs at Indiana University

[9]. NaradaBrokering has demonstrated applicability in real-time environments involving audio/video streaming where support for low-latency communications is critical [10]. It has been previously applied to streaming sensor-based observing systems, notably the NASA funded QuakeSim project [11].

NaradaBrokering acts as an overlay network, efficiently supporting the routing and dissemination of messages through a distributed, hierarchically arranged broker network. This approach enables scalability and reliability, without a dependence on the availability of a particular broker node. Additionally, NaradaBrokering provides support for a variety of transport methods including: TCP, UDP, Multicast, SSL, HTTP and ParallelTCP. Transport flexibility simplifies the creation of broker networks across administrative domains, where firewalls may impact conventional connectivity. NaradaBrokering does not impose any restrictions on the type or size of its message contents. It can even provide fragmentation/coalescing, (de)compression and parallel transport capabilities for large payloads. Content-based subscription/advertisement is supported along with the capability for XPath and regular expression based constraints for subscriptions.

Observation Data Management

We build on the capabilities of our messaging middleware to provide functionality specific to streaming observation data management. We introduce software to address the movement of observation data onto and off of the messaging substrate. This includes defining the network representation for data/metadata and supporting the subsetting of observation feeds based on powerful query tools. Core functionality is provided by a library that defines the model and format for observation data. This library provides the basis for development of publishing and subscribing entities. A typical configuration is shown in Figure 4.2. Any number of possible publishing and subscribing applications may choose to participate with a given messaging provider. Applications can be loosely grouped into three categories: sources, filters and sinks. Sources act as gateways for observation data, typically interfacing directly with instrumentation through network or serial streams. Filters subscribe to sources, perform processing or translation, then republish the data. A sink subscribes to sources or filters and acts as a destination for observation data. This discussion identifies four applications that provide the basis for a complete messaging solution. These applications require Java 1.5 or later.

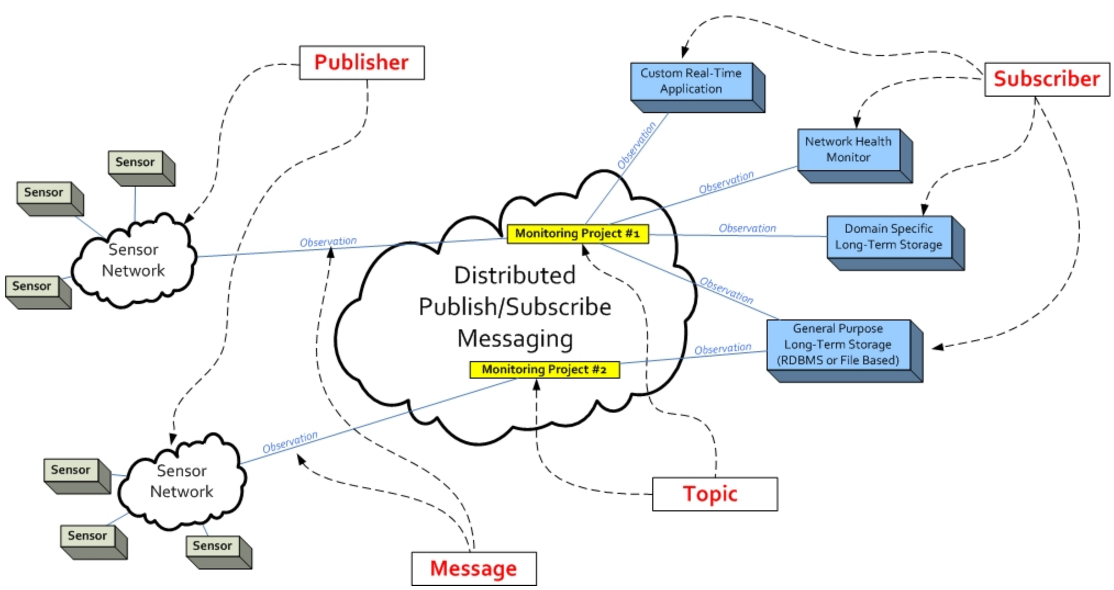


Figure 4.2: Brokered Publish/Subscribe for Environmental Observation

At its core, the observation data layer specifies a representational model of data and metadata for transmission over the messaging substrate. Our approach draws heavily from the scientific data tools developed at the Unidata Program Center. In particular, we utilize the Common Data Model (CDM) [12] component of the NetCDF-Java libraries for our native data representations. NetCDF-Java supports a multitude of scientific data models and formats including the Data Access Protocol (DAP) and the Hierarchical Data Format (HDF). Translation is performed to fit observations into a unified model that supports a variety of observation types. This library is extensible to support custom or future formats. An overview of the architecture is shown in Figure 4.3.

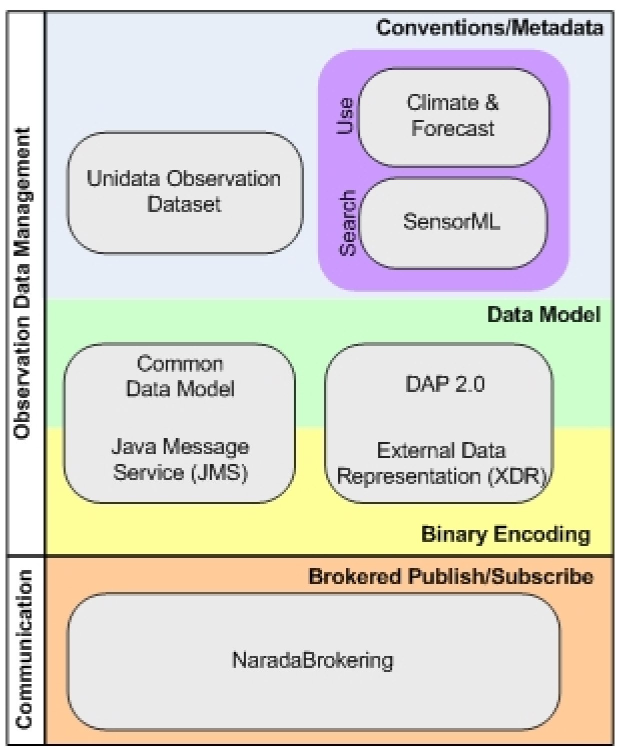


Figure 4.3: Architecture Model: Streaming Observation Data Management

The CDM identifies six scientific datatypes when classifying observation types. These datatypes correspond to different methods of data collection, e.g. grid, trajectory or station observations. An observation is a measurement or collection of measurements over time. Our current monitoring efforts have relied on stationary sensor network nodes. As a result, we have focused our development efforts on the station observation datatype. We anticipate the addition of tools specific to other scientific datatypes in future releases. By focusing our development at this level of granularity, we can provide capabilities for visualization, processing and storage of all observation stream traffic. A notable feature of CDM scientific datasets is the emphasis on the inherent spatial component of observation data. The CDM tools include support for explicit designation of geographic coordinate systems. This information is critical to interoperability with geographic information system tools.

A key characteristic of a publish/subscribe interaction is the ability for subscribers to register interest in a subset of messages. NaradaBrokering supports the use of XML and XPath to specify subscription criteria. The use of a XML document in the message header allows subscribers to identify the relevance of a message before processing its payload. In particular, the XML document may include search metadata and support functionality similar to that of a conventional metadata catalog. We explore the use of SensorML [13], specified by the Open Geospatial Consortium specification, for this purpose.

SensorML provides a means of defining the spatial and observational characteristic of a sensor or group of sensors.

Management Applications

The adoption of a native representation for observation data does not preclude the use of other compatible data models. To illustrate this, we introduce a filter application for data model translation. In the earth and atmospheric sciences, the Data Access Protocol (DAP) 2.0 specification [14] is a commonly used data model and format for network access to observation data. DAP is one of the models supported under the Common Data Model. We should note that the DAP specification is not fully supported at the messaging layer because it includes data access methods that are not applicable to our asynchronous communication style. We utilize the specification only for its designation of data model and format, although traditional DAP libraries can still be leveraged when building subscribing applications.

Source applications are responsible for publishing observation data onto the messaging substrate. Our management software includes an application that interfaces with instrumentation and sensor network nodes for data retrieval, processing and observation message creation. Two basic functionalities are supported. First, vendor specific observation streams are processed into the common data model format. Second, observation data is associated with metadata. The first functionality is an inherently complex process. As noted earlier, we incorporate a variety of sensing instrumentation, each requiring some level of customized integration. This application provides a collection of extensible classes to ease integration tasks. Additionally, common functionality e.g. TCP/IP and serial stream processing, is provided as well as messaging connectivity and broker discovery. Once the observation data and its associated metadata are fit to the CDM model, it can be published to the messaging substrate. With advances in embedded computing, it is conceivable that sensor network nodes may eventually support integration directly into the messaging substrate.

Traditional data analysis and visualization tools are not typically capable of working directly with streaming data sources. Data access is predominately file based access, either local or network. To support persistent storage of observation data, we provide a sink application for data archiving. This sink subscribes to observation messages, aggregates them into a collection of station observations and writes them to a NetCDF-3 file. The NetCDF-3 file format is widely supported by data analysis and visualization software and has API libraries in nearly every popular development environment. Additionally, NetCDF-3 files can be directly tied into the THREDDS Data Server (TDS) [15], a powerful cataloging and data distribution tool.

Access to real-time streaming observation data offers powerful opportunities for analysis and visualization. We incorporate a network health monitoring utility that monitors active streams of observation data. This application supports a web-based client developed using the Google Web

Toolkit. It utilizes AJAX and Servlet technologies to provide an interactive tool for tracking and summarizing active observation streams. The browser view, shown in Figure 4.4, has three basic components; current traffic statistics chart, spatial visualization of station locations and a table view listing each reporting station. Spatial visualization is supported through Google Maps. The table view supports a drill-down capability allowing the user to view details specific to a station, included current and recent observation data streams.

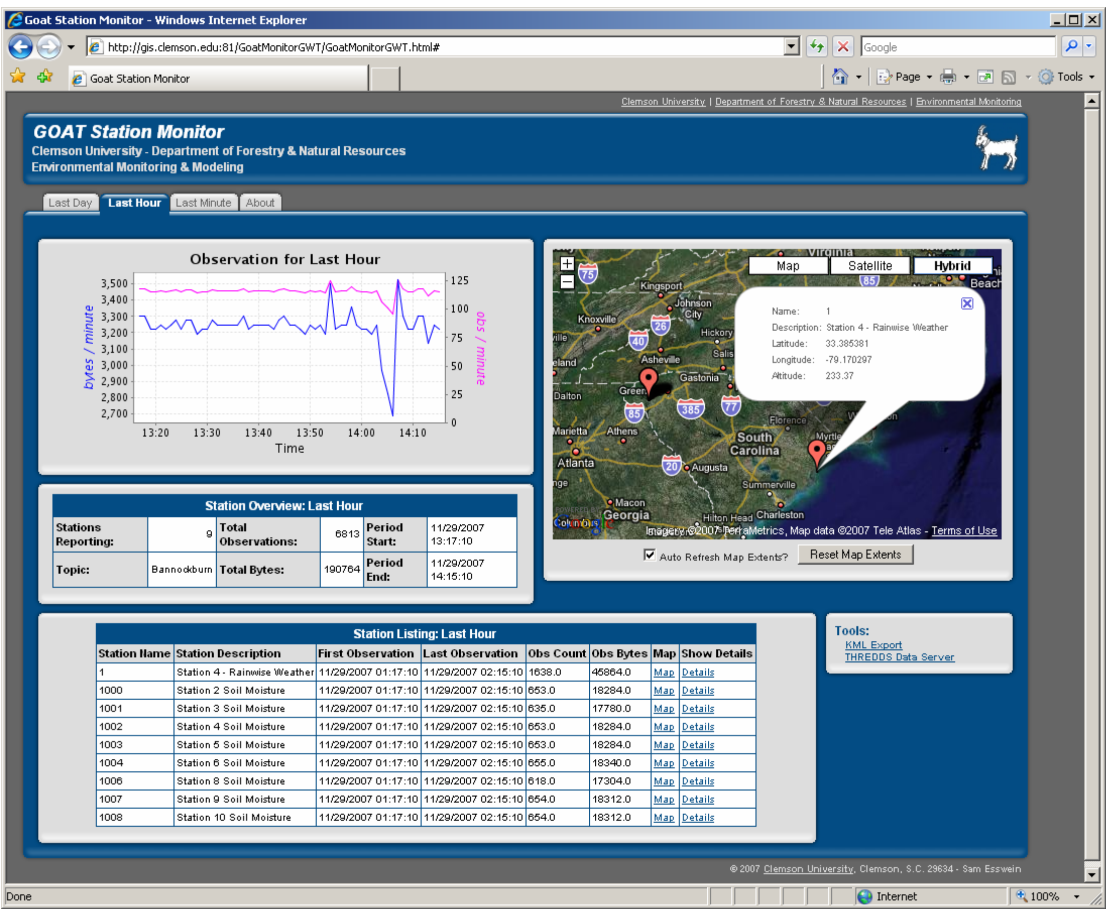


Figure 4.4: Web Application for Monitoring State of Streaming Observation Data System

Discussion

The land use change monitoring project offered the first opportunity to evaluate our messaging solution with observation data derived from a field installation. The first phase of sensing equipment installation is complete. Initial system implementation is complete with a distributed broker network and a collection of application servers. Data products are currently available to our research team via direct broker connections, a THREDDS data server and through our network health monitoring application. Thus far, our system has shown suitability for this class of streaming data management.

Currently, our streaming observation management system is implemented as a five server configuration, with three machines acting as NaradaBrokering brokers. The NaradaBrokering nodes are 2.0 GHz dual-processor Xeon machines with 1GB of RAM. Network between nodes is conventional gigabit Ethernet. Our wireless sensor network connects to a dedicated server running the observation publishing application. This connection is via a 115kbps cellular link. The network health monitor and archiving application run on a separate server under Apache Tomcat version 5.5.

The current traffic results in modest loads on the streaming observation management component. The majority of observation types are relatively small in size, generally consisting of payload sizes under 50bytes. In its current configuration, traffic originating from the deployed sensor network yields approximately one hundred observations per minute. This volume of traffic is easily within the capability of NaradaBrokering. We hope to incorporate several additional monitoring efforts in the near future. Future plans for include the addition of remote cameras for habitat monitoring, which will significantly increase traffic volume.

Related Work

We have identified three middleware offerings that provide both message-based communications for managing observation data streams. These alternatives differ from our approach in two fundamental ways. First, a specialized messaging layer provides the transmission of data streams versus our general purpose messaging substrate. Second, data is treated as an opaque object, i.e. message semantics are unknown. The first two are based around the concept of distributed ring-buffers, acting as observation repositories. Clients interact with the ring-buffer nodes to receive streaming observation data. The Ring Buffer Network Bus (RBNB) DataTurbine [16] project has a history of supporting large environmental observatories [2] [17] [18]. From an interaction perspective, DataTurbine provides time and synchronization decoupling. The second ring-buffer based solution is a commercial software package called Antelope [19] developed by Boulder Real Timer Technologies. Antelope is similar in nature to DataTurbine and has been applied to sensor-based observing systems, e.g. seismic research [19]. The third solution is the Local Data Manager (LDM) [20] software package developed by the Unidata group. LDM supports time, synchronization and a degree of space decoupling. Space decoupling is limited by thirty-two predefined types for data stream description [21]. LDM does not support user-defined or content-based subscriptions. Like DataTurbine, LDM uses static routing between messaging network nodes. This impacts reliability and increases the complexity, particularly as the number of participating nodes increases.

Future Work

Supporting web services at the streaming data management level has potential to broaden its applicability. At the messaging level, WS-Notification [22] and WS-Eventing [23] provide publish/subscribe capabilities. NaradaBrokering provides support for the WS-Eventing specification. At the data management layer, supporting a web services approach has some benefits, notably simplifying integration into larger workflows and widening support for various client platforms. XML has excelled in certain areas of environmental observation, particularly with the representation of metadata. On the other hand, the use of XML for binary data is problematic due to its inefficiency and overhead. Typically this means either embedding binary data or providing a callback reference. Both methods have their drawbacks. The draft specification for DAP 4.0 [24] may offer future opportunity to support web services at the data management level.

Future sensor network implementations will utilize intelligent sensing devices (motes) for remote data acquisition. The intent is to incorporate these devices directly into the messaging substrate. Responsibility to provide data processing and metadata association will fall to the device. This has the potential to simplify sensor deployments, sensors may even generate their own metadata, i.e. use GPS to determine sampling locations. Additionally, intelligent sensing devices may be consumers as well as publishers of observation data. This supports features like adaptive sampling parameters based on the state of other sensors.

Conclusion

We presented a streaming data management solution for environmental observing systems. Our land use change monitoring effort demonstrates the integration of this solution into a complete environmental observing system. This solution demonstrated support for heterogeneous instrumentation and support for a variety of consuming applications. We characterized various forms of communication interactions based on their decoupling dimensions. This characterization provides the support for our selection of a brokered publish/subscribe approach. We utilized the NaradaBrokering distributed messaging software to support the transmission of observation data streams. We then introduced a complete messaging solution that leverages this messaging substrate to support the requirements of a sensor-based observing system. This solution provides support for processing of observation data onto and off of the messaging substrate as well as providing data translation capabilities.

References

[1] Anonymous (2008). Geosciences Network. [Online]. Available: http://www.geongrid.org/

[2] Anonymous (2008). National Ecological Observatory Network. [Online]. Available: http://www.neoninc.org/

[3] Anonymous (2008). Water and Environmental Research Systems Network (WATERS). [Online]. Available: http://www.watersnet.org/

[4] William Conner. (2008). Program of Integrated Study for Coastal Environmental Sustainability. [Online]. Available: http://www.clemson.edu/baruch/pisces/

[5] Anonymous (2008). Maxstream Radio Modules by Digi International. [Online]. Available: http://www.maxstream.net

[6] P. T. Eugster, P. A. Felber, R. Guerraoui and A. Kermarrec, "The Many Faces of Publish/Subscribe," *ACM Computing Surveys,* vol. 35, pp. 114-131, 2003.

[7] E. Gamma, R. Helm, R. Johnson and J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software.* Reading, Mass.: Addison-Wesley, 1995, pp. 395.

[8] I. Foster, C. Kesselman, J. Nick and S. Tuecke, "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration," *Open Grid Services WG, Global Grid Forum,* June 22, 2002.

[9] Anonymous (2008). The NaradaBrokering Project @ Indiana University. [Online]. Available: http://www.naradabrokering.org/index.html

[10] G. Fox, G. Aydin, H. Bulut, H. Gadgil, S. Pallickara, M. Pierce and W. Wu, "Management of real-time streaming data Grid services: Research Articles," *Concurr. Comput. : Pract. Exper.,* vol. 19, pp. 983-998, 2007.

[11] Anonymous (2008). NASA - QuakeSim. [Online]. Available: http://quakesim.jpl.nasa.gov/

[12] Anonymous (2008). Common Data Model (CDM) - Unidata. [Online]. Available: http://www.unidata.ucar.edu/software/netcdf/CDM/

[13] M. Botts and A. Robin. OpenGIS sensor model language (SensorML) implementation specification. [Online]. Available: http://portal.opengeospatial.org/files/?artifact_id=21273

[14] J. Gallagher, N. Potter, T. Sgouros, S. Hankin and G. Flierl. The data access protocol - DAP 2.0. [Online]. Available: http://www.opendap.org/pdf/ESE-RFC-004v1.1.pdf

[15] Anonymous Thematic Realtime Environmental Distributed Data Services (THREDDS). [Online]. Available: http://www.unidata.ucar.edu/projects/THREDDS/

[16] Anonymous (2008). Open Source DataTurbine Initiative. [Online]. Available: http://www.dataturbine.org/

[17] Anonymous (2008). Consortium of Universities for the Advancement of Hydrologic Science. [Online]. Available: http://www.cuahsi.org/

[18] Anonymous NEES. [Online]. Available: http://www.nees.org

[19] Anonymous (2008). Boulder Real Time Technologies, Antelope System. [Online]. Available: http://www.brtt.com/

[20] Anonymous (2008). Local Data Manager (LDM) - Unidata. [Online]. Available: http://www.unidata.ucar.edu/software/ldm/

[21] R. K. Rew and A. Wilson, "The Unidata LDM System: Recent Improvements for Scalability," in 17th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology, 2001, pp. 171-174.

[22] S. Graham, D. Hull and B. Murray. Web service base notification 1.3. OASIS, [Online]. Available: http://docs.oasis-open.org/wsn/wsn-ws_base_notification-1.3-spec-os.pdf

[23] D. Box, L. F. Cabrera, C. Critchley, F. Curbera, D. Ferguson, A. Geller, S. Graham, D. Hull, G. Kakivaya and A. Lewis. Web services eventing. W3C, [Online]. Available: http://www.w3.org/Submission/2006/SUBM-WS-Eventing-20060315/

[24] J. Gallagher, N. Potter and T. Sgouros. DAP data model specification 4.0. [Online]. Available: http://www.opendap.org/pdf/dap_objects.pdf

CHAPTER FOUR

CONCLUSION

This research introduces an environmental observing system in support of a study investigating the hydrological and ecological impacts of land use change in coastal areas. This research was broken into three topics representing the three main areas of an environmental observing system. The first topic addresses the nature of the impacts of land use change study. The second focuses on the design and implementation of a wireless sensor network in support of the study. The third topic deals with the development of a novel approach for efficiently managing real time streams of environmental observation traffic. Each topic was devoted a chapter in this thesis. Taken together, these topics comprise a complete environmental observing system.

Chapter two provided an overview of an ongoing, multidisciplinary study investigating the hydrological and ecological impacts of land use change on the coast of South Carolina. This chapter focuses on the surface and groundwater hydrology components of the study. This includes an explanation of sampling and monitoring locations and methods. A justification of the purpose of the study was included, along with plans for future expansion of monitoring and sampling.

Chapter three provided an overview of the challenges faced with wireless sensor networks implementations for environmental monitoring applications. An explanation of wireless sensor network technologies is provided, including an evaluation of current offerings. A complete remote data acquisition approach is presented; including physical implementation, radio considerations including frequency, antenna and cabling decisions. Challenges include the spatial and vegetation density characteristics of the site, from both a radio and installation perspective. Also included is a proposed direction for future research based on sensing implementations utilizing devices with greater processing and sensing capabilities.

Chapter four presented a streaming data management solution for environmental observing systems. This solution is evaluated using the land use change study from chapter two. This solution demonstrated support for heterogeneous instrumentation and support for a variety of consuming applications. We characterized various forms of communication interactions based on their decoupling dimensions. This characterization provides the support for our selection of a brokered publish/subscribe approach. The NaradaBrokering distributed messaging software is used to support the transmission of observation data streams. A complete messaging solution is introduced that leverages this messaging substrate to support the requirements of a sensor-based observing system. This solution provides support for processing of observation data onto and off of the messaging substrate as well as providing data translation capabilities.

APPENDICES

Appendix A

Weather Station NCDump Output

```
netcdf RainwiseWeatherStations_2008042215.nc {
 dimensions:
   record = UNLIMITED;   // (1838 currently)
   station = 1;
   station_id_strlen = 5;
   station_description_strlen = 28;
   station_type_strlen = 25;
   time_strlen = 30;
   id_len = 5;
 variables:
   double latitude(station=1);
     :units = "degrees_north";
     :long_name = "station latitude";
   double longitude(station=1);
     :units = "degrees_east";
     :long_name = "station longitude";
   double altitude(station=1);
     :units = "feet";
     :long_name = "station altitude";
   char station_id(station=1, station_id_strlen=5);
     :long_name = "station identifier";
   char station_description(station=1, station_description_strlen=28);
     :long_name = "station description";
   char station_type(station=1, station_type_strlen=25);
     :long_name = "Identifies type of station for processing";
   int numChildren(station=1);
     :long_name = "number of children in linked list for this station";
```

```
int lastChild(station=1);
  :long_name = "record number of last child in linked list for this station";
int firstChild(station=1);
  :long_name = "record number of first child in linked list for this station";
char timeString(record=1838, time_strlen=30);
  :long_name = "ISO-8601 Date";
int time(record=1838);
  :long_name = "seconds since 1970-01-01 00 UTC";
  :units = "seconds since 1970-01-01 00 UTC";
int prevChild(record=1838);
  :long_name = "record number of previous child in linked list";
int parent_index(record=1838);
  :long_name = "index of parent station";
int nextChild(record=1838);
  :long_name = "record number of next child in linked list";
char report_id(record=1838, id_len=5);
  :long_name = "Station id";
  :_FillValue = "";
float temperature(record=1838);
  :long_name = "Temperature";
  :units = "F";
float relativeHumidity(record=1838);
  :long_name = "Relative Humidity";
  :units = "%";
float pressure(record=1838);
  :long_name = "Barometric Pressure";
  :units = "inHg";
float windSpeed(record=1838);
  :long_name = "Wind Speed";
  :units = "mph";
```

```
  float windDirection(record=1838);
    :long_name = "Wind Direction";
    :units = "degrees";
  float rainfall(record=1838);
    :long_name = "Rainfall";
    :units = "inches";
  float voltage(record=1838);
    :long_name = "Voltage";
    :units = "V";
 :Conventions = "Unidata Observation Dataset v1.0";
 :cdm_datatype = "Station";
 :title = "Bannockburn Maxstream Sensor Network";
 :desc = "None";
 :time_coordinate = "time";
 :geospatial_lat_min = "-90.0";
 :geospatial_lat_max = "90.0";
 :geospatial_lon_min = "0.0";
 :geospatial_lon_max = "360.0";
 :time_coverage_start = "2007-06-30T00:00:01Z";
 :time_coverage_end = "2009-6-30T23:59:00Z";
}
```

Appendix B

Weather Station CSV Output

date	timeString	time	prevChild	parent_index	nextChild	report_id	temperature	relativeHumidity	pressure	windSpeed	windDirection	rainfall	voltage
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:13:43 GMT	1208891623	-1	0	1	1	73	59	29.83	2	135	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:13:45 GMT	1208891625	0	0	2	1	73	59	29.83	2	135	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:13:47 GMT	1208891627	1	0	3	1	73	59	29.83	2	157	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:13:49 GMT	1208891629	2	0	4	1	73	59	29.83	3	180	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:13:51 GMT	1208891631	3	0	5	1	73	59	29.83	3	180	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	4	0	6	1	73	60	29.83	4	180	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	5	0	7	1	73	60	29.83	3	135	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	6	0	8	1	73	60	29.83	2	135	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	7	0	9	1	73	60	29.83	1	135	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	8	0	10	1	73	60	29.83	1	225	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	9	0	11	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	10	0	12	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	11	0	13	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	12	0	14	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	13	0	15	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	14	0	16	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	15	0	17	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	16	0	18	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	17	0	19	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	18	0	20	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	19	0	21	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	20	0	22	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	21	0	23	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	22	0	24	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	23	0	25	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	24	0	26	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	25	0	27	1	73	60	29.83	0	247	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	26	0	28	1	73	60	29.83	0	45	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	27	0	29	1	73	60	29.83	0	360	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	28	0	30	1	73	60	29.83	0	315	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	29	0	31	1	73	60	29.83	0	315	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	30	0	32	1	73	60	29.83	5	315	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	31	0	33	1	73	60	29.83	6	315	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	32	0	34	1	73	60	29.83	6	315	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	33	0	35	1	73	60	29.83	9	337	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	34	0	36	1	73	58	29.83	7	360	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:55 GMT	1208891695	35	0	37	1	73	58	29.83	4	360	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:56 GMT	1208891696	36	0	38	1	73	58	29.83	2	360	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:14:58 GMT	1208891698	37	0	39	1	73	58	29.83	1	360	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:15:00 GMT	1208891700	38	0	40	1	73	58	29.83	0	360	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:15:02 GMT	1208891702	39	0	41	1	73	58	29.83	0	360	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:15:04 GMT	1208891704	40	0	42	1	73	58	29.83	0	360	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:15:06 GMT	1208891706	41	0	43	1	73	58	29.83	0	22	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:15:08 GMT	1208891708	42	0	44	1	73	58	29.83	0	112	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:15:10 GMT	1208891710	43	0	45	1	73	58	29.83	2	112	2	0.24
Tue Apr 22 15:14:08 EDT 2008	22 Apr 2008 19:15:12 GMT	1208891712	44	0	46	1	73	58	29.83	4	112	2	0.24
Tue Apr 22 15:16:16 EDT 2008	22 Apr 2008 19:15:15 GMT	1208891715	45	0	47	1	73	58	29.83	5	112	2	0.24
Tue Apr 22 15:16:16 EDT 2008	22 Apr 2008 19:15:17 GMT	1208891717	46	0	48	1	73	58	29.83	5	112	2	0.24