Large Scale Wireless Network Systems: Experience, Observations, and Theories

Xiang-Yang Li

Department of Computer Science
Illinois Institute of Technology
www.cs.iit.edu/~xli xli@cs.iit.edu
Tsinghua University, EMC Visiting Chair Professor
Wireless Sensor/Actuator Networks

Bridging the digital world and physical world

Information

Sensed Data

Physical Environment

Sensed Data

Sensed Data

Sensed Data

Sensed Data

Sensed Data

Sensed Data
Wide Applications: CPS, IOT

- Environment
- Transportation
- Smart Grid
- Security
- Green Building
- Agriculture
- Industry Monitoring
- Logistic and Supply Chain
- Health Care
- CPS/IoT
Why large scale wireless network?

- Scalability
- Diversity (spatial, temporal)
- Asymptotical Behavior
- Application Requirement
Presentation Outline

OceanSense  GreenOrbs  CitySee

Networking observations

ZIMO: Coexistence  Capacity: the Limit
Real World Systems


2. GreenOrbs (2009-)

3. CitySee (2011-)
OceanSense

The Hong Kong University of Science and Technology

Illinois Institute of Technology

Tsinghua University
Motivation

- Silt Deposition problem of Qingdao Port:
  - Qingdao port:
    - one of the ten busiest ports in the world
  - Silt Deposition:
    - Affect the water depth
    - High uncertainty and high instant uncertainty (tide, wind, etc.)
Monitor the sea!

- The **first** sea environment monitoring sensor network system in China
- More than **120** sensor nodes
- Temperature, Light, Sea depth
Deployment

• Deployed in the Yellow sea near Qingdao, China
GreenOrbs

http://www.greenorbs.org/
Motivation

- Carbon sequestration
- Study on biodiversity
- Canopy closure estimates
- Fire risk evaluation
GreenOrbs

- Go to the wild!
  - Supporting forestry research and applications
  - Multiple deployments, each > 330 sensor nodes
  - Temperature, Light, CO2
## Deployment: Overview

<table>
<thead>
<tr>
<th>Place</th>
<th>Area</th>
<th>Duration</th>
<th>Battery</th>
<th>Scale</th>
<th>Network Diameter</th>
<th>Duty Cycle</th>
<th>Data Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>University woodland #1</td>
<td>20,000 m²</td>
<td>1 month (2008)</td>
<td>800 mAh 1.5V</td>
<td>50</td>
<td>6 hops</td>
<td>No</td>
<td>15 Mbytes</td>
</tr>
<tr>
<td>University woodland #2</td>
<td>20,000 m²</td>
<td>10 months (2009)</td>
<td>2200 mAh 1.2V</td>
<td>120</td>
<td>10 hops</td>
<td>5%</td>
<td>272 Mbytes</td>
</tr>
<tr>
<td>University woodland #2 and #3</td>
<td>40,000 m²</td>
<td>1 year (2009.12~)</td>
<td>~8000mAh, 1.5V</td>
<td>330</td>
<td>12 hops</td>
<td>8% or No</td>
<td>140 Mbytes</td>
</tr>
<tr>
<td>Tianmu Mountain</td>
<td>200,000 m²</td>
<td>1.5 months (2009)</td>
<td>~8000mAh, 1.5V</td>
<td>50</td>
<td>10 hops</td>
<td>5%</td>
<td>3 Mbytes</td>
</tr>
<tr>
<td>Tianmu Mountain</td>
<td>200,000 m²</td>
<td>1.5 year (2009.10~)</td>
<td>~8000mAh, 1.5V</td>
<td>200</td>
<td>~ 20 hops</td>
<td>5%</td>
<td>10 Mbytes</td>
</tr>
</tbody>
</table>

### Notes
- ~ 20 hops
- 12 hops
- 8% or No
Deployment: Nodes in the Wild
CitySee
City-Wide Urban Sensing
Motivation: Global Warming

- Starting from Global Climate Changes
  - Emission of large volume of greenhouse gases is the main reason for global warming
    - CO2, N2O, CH4, HFCs, PFCs, SF6
  - The most greenhouse gases is CO2
  - CO2 generation of human activities: in the city
CitySee

- Back in the city!
  - Large scale indoor/outdoor environment monitoring
  - More than 1200 sensor nodes
  - Temperature, Light, CO2
  - Mesh routers
System Architecture

Data Center and Apps

MESH Networks

WSN
Deployment: Locations

Cover more than 1.2 KM² urban area of the Wuxi City

- Thermal Power Plant
- Water Source
- High emission Factories
- Residential Area
- Development Zone
- Railway Station
System Deployment
## Deployment: Nodes Deployed

<table>
<thead>
<tr>
<th></th>
<th>Normal node</th>
<th>Carbon node</th>
<th>Mesh node</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Microcontroller</strong></td>
<td>MSP430f1611</td>
<td>MSP430f1611</td>
<td>ARM7</td>
</tr>
<tr>
<td><strong>Type of sensor reading</strong></td>
<td>Temperature, humidity, light.</td>
<td>CO₂</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Radio module</strong></td>
<td>IEEE 802.15.4 CC2420 2.4GHz</td>
<td>IEEE 802.15.4 CC2420 2.4GHz</td>
<td>IEEE 802.11b NetCard 5.8GHz</td>
</tr>
<tr>
<td><strong>Communication range (m)</strong></td>
<td>150~200</td>
<td>150~200</td>
<td>5000~6000</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td>2 AA batteries (3V)</td>
<td>12V Rechargeable battery</td>
<td>110V~220V AC</td>
</tr>
<tr>
<td><strong>Power consumption - sleeping (mW)</strong></td>
<td>0.6~1.2</td>
<td>2.4~4.8</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Power consumption - sensing (mW)</strong></td>
<td>60~90</td>
<td>~2160</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Power consumption - communication (mW)</strong></td>
<td>60~90</td>
<td>60~90</td>
<td>3000~25000</td>
</tr>
<tr>
<td><strong>Manufactory cost (USD)</strong></td>
<td>~80</td>
<td>~260</td>
<td>~800</td>
</tr>
</tbody>
</table>
Presentation Outline

OceanSense  GreenOrbs  CitySee

Networking observations

ZIMO: Coexistence

Capacity: the Limit
LESSONS
Lesson 1

- System that work in labs fails horribly in practice
  - OceanSense:
    - System run out of battery in a week (labs run in months)
    - Nodes destroyed by water
    - Devices stolen by people: they are interested in the sticks!
  - GreenOrbs:
    - Nodes destroyed by flooding
  - CitySee:
    - Installation needs the coordination of various government departments
    - Require nice encapsulation
Lesson 2

**Encapsulation? Encapsulation!**
- Solutions to many of the previously mentioned problems
- **OceanSense:**
  - Waterproof, considering factors such as tide, wind, etc.
- **GreenOrbs:**
  - Waterproof, allow accurate collection of humidity and luminosity
- **CitySee:**
  - Made the nodes nice-looking!
Lesson 3

- Need good visualization tools
  - Allow diagnostics
  - Easy to interpret the data and locate the problems
Lesson 4

- Deployment
  - Balance between accuracy, coverage, sustainability and cost
  - Regions that doesn’t allow deployment
    - Not allowed by the nature (physically infeasible)
    - Not allowed due to bad signal (interference, obstruction, etc)
    - Forbidden by the government
  - Deployment challenges: bamboo, large trees
Lesson 5

- Control the cost!
  - Cost reduction is a must when you need so many nodes
    - Node cost
    - Labor cost
NETWORK OBSERVATIONS: GREENORBS
Traffic distribution: balanced in CTP?

5% nodes account 80% traffic.

90% nodes have very low traffic.

The traffic distribution is relatively stable over time.
Causes of Packet Losses

- Packet Delivery Ratio (PDR) about 85%
  - Link loss (61%) vs. Node drops (39%)
- Faulty behavior on forwarding nodes

Cumulative distribution of packet loss

Causes of packet drops on sensor nodes
Packet Loss Diagnosis

December 10, 2010; 400 nodes, 60,000m²

Data of 10 days:
1,137,430 packets received
181,862 packets lost

- The green nodes with PRR > 90%.
- The red nodes with PRR < 90%,
- The radius indicates the number of lost packets
Packet Losses: Non-ACK

- 84,030 packet loss due to non-ack
  - 46.2% of total losses
  - 68,444 caused by physical environment (bad links)
Packet Losses: Non-ACK

- 84,030 packet loss due to non-ack
  - 46.2% of total losses
  - 4,361 caused by interferences (contention <--reboot, loop)
Packet Losses: Corrupted Packets

- 9,511 corrupted packets
  - 9037 real losses (after consider retransmission)
  - ~ 5% of total loss
Packet Losses: Routing Loop

- 5,178 packet loss due to overflow from routing loop
  - 2.9% of total losses
  - 93% of overflow events did not result in packet loss
## Packets Loss Summary

<table>
<thead>
<tr>
<th>Root cause</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. sink-side failure</td>
<td>12.5%</td>
</tr>
<tr>
<td>1.1 vertical banding</td>
<td>12.45%</td>
</tr>
<tr>
<td>2. corruption</td>
<td>5%</td>
</tr>
<tr>
<td>3. overflow drops</td>
<td>2.87%</td>
</tr>
<tr>
<td>3.1 loop overflow drops</td>
<td>2.85%</td>
</tr>
<tr>
<td>3.2 non-loop overflow drops</td>
<td>0.02%</td>
</tr>
<tr>
<td>4. no-ack drops</td>
<td>46.2%</td>
</tr>
<tr>
<td>4.1 env-no-ack drops</td>
<td>37.6%</td>
</tr>
<tr>
<td>4.2 interference-no-ack drops</td>
<td>2.4%</td>
</tr>
<tr>
<td>5. reboot (direct impact on loss)</td>
<td>~0</td>
</tr>
</tbody>
</table>

About **35%** packet losses are unidentified now.
NETWORK OBSERVATIONS: CITYSEE
Where Packets are Lost?

From figure (b)(c)(d), the load distribution is closely related with the spatial property of the deployment.

Is there any node suffering from heavy packet drops?

Where Packets are Lost?

(a) # of packets at sink
(b) # of packets transmitted
(c) Radio duty cycle
(d) # of task executions

From figure (a), no apparent correlation between packet drop and the node location.

Does each node play the same role in the network?
Traffic Distribution

• Small portion of “critical nodes”, verifies the same finding observed from GreenOrbs

• Traffic dynamics exhibits different pattern, e.g. burstness on some nodes
Nodes closer to the sink have a more stable topology than nodes that are far away.
Summary

Many challenges to make it

1. **Sustainable** --- energy efficiency and fault diagnosis?

2. **Robust** --- co-existence?

3. **Scalable** --- large scale performance?

4. **Predictable** ---- under varying environment?
Presentation Outline

OceanSense  GreenOrbs  CitySee

Networking observations

ZIMO: Coexistence  Capacity: the Limit
ZIMO: CROSS-TECHNOLOGY MIMO TO HARMONIZE COEXISTENCE OF ZIGBEE WITH WIFI
Coexistence in ISM Band

ISM band interferences are pervasive and crowded. WiFi signal is the primary and first class passenger.

- Existing Works mainly protect WiFi signal and mitigate cross technology interference [TIMO]
- Some ZigBee signal protection works need modifications or degrade WiFi [Sensys10, Liang]
Experiment Setup

- Two ZigBee Nodes (TX&RX)
- WiFi APs are in IIT campus
- ZigBee nodes are configured to receive full spectrum interference in full time scale
- Adding controllable AP for tunable interference
Effect of WiFi interference on ZigBee

- Short and frequent WiFi data transmission (i.e., flash) play the main role of WiFi interference on ZigBee.

- Power-law like distribution indicates the shorter flashes interfere ZigBee signal with exponentially increasing probability, which is a drastic threat for ZigBee signal.
The WiFi interference is distributed across ZigBee symbols, rather than concentrated on particular positions. We need to resort to the signal processing techniques for fundamental solutions.
ZIMO: Sink Based Design

Cons No.1: ZIMO has more antennas than WiFi AP (N+1)
Cons No.2: ZIMO needs at least one preamble is clear
Cons No.3: Can work with one ZigBee and multiple WiFi
How ZIMO works?

Frame Detection & Identification

ZigBee Nullification

ZigBee Decoding

Interference Nullification

WiFi Nullification

WiFi Decoding

Accurate WiFi Signal Recovery

Cross Technology Interference Cancellation

ZigBee Decoding

WiFi Signal Cancellation

Y'
Main challenges

- **Channel Coefficient Recovery**
  - Interference in frequency domain
  - Sufficient for decoding, insufficient for accurate signal recovery

- **CFO Compensation**
  - Well done for preamble, insufficient for whole data scale
  - Extremely large with the increasing packet length
Where are opportunities?

- **Frequency** domain is partially overlapped
- **Time** domain also partially overlapped
- **Power** domain shows significant difference
Where are opportunities?

For channel coefficient:
Interpolation is simple, effective
Where are opportunities?

Linear regression is accurate enough for CFO
Implementation

- Implement using USRP2 N200
  - IEEE STD 802.15.4, 2 MHz Bandwidth
  - OFDM is 20 MHz Bandwidth
- Real trace driven ZIMO decoding
- No carrier sense and MAC timing control
Experimental Results: Macro Benchmark

- Recovery ratio

![Graph showing recovery ratio vs SNR (dB)](image)

![CDF plot with Intact WiFi Preamble, Intact ZigBee Preamble, and WiFi Addon Packets](image)
Experimental Results: Macro Benchmark

- Throughput

1. Zigbee Baseline
2. WiFi Baseline
3. Interference patterns
Asymptotical Capacity
Two Capacity metrics - channel

- Channel capacity
  - achievable single-hop data rate

\[ C = \log(1 + SINR) \]

Shannon channel theory

ACK
Two Capacity metrics - transport

- transport capacity
  - end-to-end multi-hop throughput

Capacity
Impact factors:

- Network Size
- Networking Models
- Inference Models
- Traffic Models
Various Models

- **deployment models**
  - arbitrary networks
  - random networks

- **network scaling models**
  - dense networks
  - extended networks

- **Communication (Interference) models**
  - the protocol model (PrIM)
  - Fixed Range Protocol Model (fPrIM)
  - physical model (PhIM)
  - generalized physical model (GphM, also called GCM)
  - Others

- **Traffic models**
  - Unicast
  - Broadcast
  - Multicast
  - Anycast
  - Many-to-one

**Fixed Range Protocol Interference Model (fPrIM)**
- Link rate \( w \) bps
- Transmission range \( r \)
- Interference range \( R \)
- Receiver \( v \) should not be interfered by other senders

Idealistic, but give us a reasonable scenario to study

**Gaussian Channel Model**
Results Summary
**Milestone Results**: Unicast, PrIM

- **Aggregated capacity**: scales $n$

  - Grossglauser and Tse, 2002, *With mobility & large delay*
  
  - AN, Gupta & Kumar, 2000
  
  - RDN,

  - Li, Goeckel and Towsley, 2006, RDN
    - Network Coding does NOT Matter

  - Alireza Keshavarz-Haddad et al, 2014, AN
    - Capacity gain of NC $\leq \pi$ for all IM

- **Thomas Moscibroda, 2007**, *Worst-Case Deployment, no BC*

- **Upper Bound**
- **Lower Bound**
- **Overlapping Bounds**
Milestone Results: **Unicast, PhIM**

- **Aggregated capacity**

  - $\sqrt{n}$
  - $\frac{n}{\log n}$

- **Franceschetti et al., 2007, AN, RDN**
  - Gap in [Gupta] can be closed

- **AN, Gupta & Kumar, 2000**
  - RDN, Gap exists
Milestone Results: **Broadcast, PrIM**

A. Keshavarz-Haddad et al, 2006, RDN

- nodes number, radio range, the area, and mobility do not matter

Aggregated capacity

\[ \sqrt{n} \]

\[ \sqrt{\log n} \]

\[ \frac{1}{\sqrt{(\log n)^\alpha}} \]
Milestone Results: **Multicast, PrIM**

\[ \text{Capacity}, \; \land (n, n_d) \]

\[ \sqrt{n} \]

\[ \sqrt[n]{\log n} \]

\[ \frac{\sqrt{n}}{\sqrt[nd]{\log n}} \]

0 1 0 \quad n_d \quad n

\[ n_s = n^\epsilon, \; n_s n_d = \Theta(n), \; \epsilon \in (0,1] \]

X. Shakkottai et al, Mobihoc, 2007, **RDN**

**Upper Bound**

\[ n_s = n^\epsilon, \; n_s n_d = \Theta(n), \; \epsilon \in (0,1] \]
Our Results: Multicast, PrIM

Li et al, MobiCom 2007, REN, \( n_s = \Theta(n) \)
Brief Summary

- The aggregate multicast capacity of $n$ sessions is

$$\Lambda_{n_d}(n) = \begin{cases} 
\Theta(\sqrt{\frac{n}{\log n}} \cdot \frac{W}{\sqrt{n_d}}) & \text{when } n_d = O\left(\frac{n}{\log n}\right) \\
\Theta(W) & \text{when } n_d = \Omega\left(\frac{n}{\log n}\right)
\end{cases}$$

- Our results unify previous results

1. **Unicast** (when $n_d=2$): $\Theta(\sqrt{\frac{n}{\log n}} \cdot W)$ by Gupta and Kumar

2. **Broadcast** (when $n_d=n$): $\Theta(W)$ by Keshavarz-Haddad et al., Mobicom’06

3. **Multicast** ($n_s=n^\epsilon$ and $n_d=n^{1-\epsilon}$), $O\left(\sqrt{\frac{n}{n_d \log n}}\right)$ by Shakkottai et al., Mobihoc’07
Our Results: Multicast, GCM

Multicast Capacity for REN, \( n_s = \Theta(n) \),
Li et al. [MobiCom 2008]. Wang, Li et al. [INFOCOM 2010, 2011].
Protocol Interference Model

General Approaches
Multicast under Protocol Model

- **Data Copies Argument** (upper bound)
  - Estimate the expected (or asymptotic lower bound) number of nodes $N(b)$ that received (or listened) a bit $b$.
  - Capacity at most $n \cdot W / N(b)$
    - since all nodes receive at rate at most $n \cdot W$. 

![Diagram of multicast network](image)
Upper-bound Proof Flow

Capacity Upper-Bound $\frac{nW}{N(b)}$

Data Copies Lower-Bound $N(b) \geq \frac{\tau \sqrt{n_d a \cdot n}}{2c_0 a}$

Density $\frac{n}{a^2}$ w.h.p.

Edge Length of $T \geq \eta_T \sqrt{n_d a}$ w.h.p.

Area $D(T) = \Theta(\|T\| \cdot r)$

Area $D(T)$ of active nodes $\geq \frac{\tau \sqrt{n_d a r}}{c_0}$ w.h.p.

$\|T\| \geq \eta \|EMST\|$ w.h.p.

Length of $EMST \geq \tau \sqrt{n_d a}$ w.h.p.
Lower-Bound: Routing and Scheduling

- **Build EMST**
  - Routing structure using EMST as backbone
  - Need to bound the conflict and total data copies
    - The lower-bound of multicast tree length w.h.p.? EMST?
    - Maximum number conflicting flows in the network w.h.p
      - Using VC dimension (proved to be $O(\log n_d)$), and VC theorem
Lower-bound Proof Flow

- Capacity Lower-Bound: $\frac{nW}{\Delta \cdot N(b)}$
- Data Copies Upper-Bound: $N(b)$
- Area $D(T)$ of active nodes $\leq$
- Density $\leq 2 \frac{n}{a^2}$ w.h.p.
- Edge Length of $T$ $\leq$
- Area $D(T) = \Theta(||T|| \cdot r)$
- Length of $EMST \leq$
- $||T|| \leq c_2 ||EMST||$
Gaussian Channel Model

General Approaches
Multicast under Gaussian Model

- Two kind of links
  - Inside Links
  - Outside Links
Relationship between links

- $l_c$: max link length in giant component.
- $\bar{l}_c$: the max distance between any node *not in GC and the giant cluster*

If $l_c = o(\sqrt{\log n / \zeta})$ then $\zeta \cdot l_c \cdot \bar{l}_c = \Omega(\log n)$
There is a link $uv$, *that will be used by many flows* (say $f$) ⇒ *the minimum data rate*

- $\min \lambda_i \leq \text{rate supported by } uv / f$
There is an isolated cluster $C$ of nodes, and $f$ flows will have links going inside this cluster.

- $\min \lambda_i \leq \text{total rate supported by links reaching } C/f$
Lower Bounds Techniques

- Highway systems
  - Cell is of $O(1)$ nodes inside
  - from *percolation* theory
  - First used by Tse et al
Low Bound: Routing, Scheduling

- First build EMST of receivers
- Build **highway** using cell size 1
  - Each highway link data rate $O(1)$
- Build **second-class highway** using cell size $(\log n)^{1/2}$
- Node sends its data to highway (solid lines) by **multi-hop** second class highway (dashed line)
Our New Techniques

- **Parallel Arterial Road Systems**
  - Longer links to connect isolated nodes to highway

- **Parallel Scheduling**

![Diagram](image)
Other Research
Cyber Physical Systems
Cognitive Radio Networks
Our iGaze Glasses
TAGORAM

Drawing in the Air

30cm

40cm

UHF Passive Tag

UHF Antenna

3m
Acknowledgments
PhD Students (alumni 10)

UNCC  GSU  financial  Google  W. Oregon

Tsinghua  financial  Motorola  UT Dallas  Toledo
Current PhD Students
MS Students
Students in IIT
Domestic Students
Deployment Videos
Thank you!

Xiang-Yang Li (李向阳)
Professor, IIT, USA
www.cs.iit.edu/~xli
xli@cs.iit.edu