ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Myths and Realities of Sensor Network Data Management

Gustavo Alonso
Institute of Pervasive Computing
Department of Computer Science
Swiss Federal Institute of Technology (ETH Zurich)
alonso@inf.ethz.ch - www.iks.inf.ethz.ch

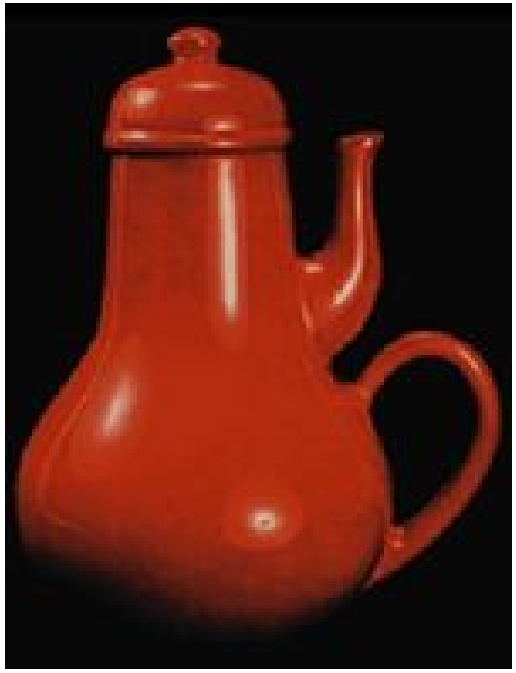
DMSN'07 – Vienna, Austria

About me ...

Experience with sensor networks

- National Competence Center for Research in Mobile Information and Communication Systems (NCCR MICS)
- Sientific data management
  - HEDC = TB repository for solar astrophysics (VLDB 2002, SIGMOD 2003)
  - BioOpera/JOpera = Protoemics data processing (ICDE 2001, ...)
- Sensor networks
  - Multiquery optimization (MASS 2006)
  - SwissQM Virtual Machine (CIDR 2007, EuroSys 2007)
  - SwissGate ...
- Interest not so much on concrete algorithms or point problems but on the system as a whole

Sensor networks:
From Research Dreams to Real Problems



Taking sensor networks from the lab to the jungle

Andres S. Tanenbaum, Chandana Gamage, and Bruno Crispo
IEEE Computer, August 2006

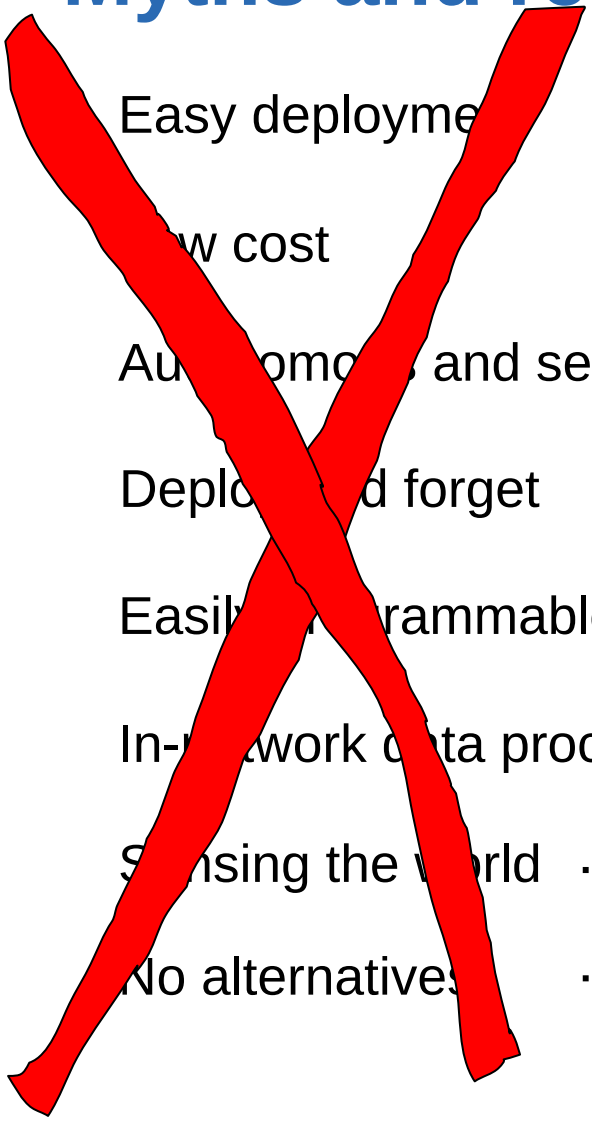
Researchers must address the systems aspects of Wireless Sensor Networks

Outline of the talk

- **Reality Check = The role of sensor networks today**
- **Challenges and opportunities**
  - Easily programmable
  - Easy deployment
  - In-network data processing
  - Sensing the world
- **The road ahead**
  - Large scale, long term deployments
  - Linking to the IT infrastructure
- **Conclusions**

Reality Check

Myths and realities of sensor networks



Easy deployme	Not today
w cost	Not today
Au omo and self organized	In the lab
Deplo d forget	In the lab
Easil rammable	Bad joke
In- twork d ta processing	Research toy
S nsing the w rld	Hammer and nail …
No alternative	Not true

The Role of sensor networks today

- **The early vision**
  - Smart dust
    - Very many small, cheap devices
  - Large scale pervasive monitoring
    - Border control
    - Throw from an airplane – gather data
- **The reality today:**
  - (low resolution, limited) Scientific instrumentation
    - Not many devices, typically expensive
    - Gather data and store – send using GSM, WiFi
  - Industrial applications
    - Different requirements / constraints

What does this mean?

Realistic WSNs today imply:

- Minimizing multi-hop routes (one hop best)
  - Hierarchical networks
  - Different nodes for different functions
- No or very minimal in-network data processing
  - Difficult to program
  - All data needed anyway
- Value added no longer in the WSN
  - Sensors and storage
  - Cost of deployment and equipment significant
- Hybrid communications
  - Mobile phones / SMS
- Power not always an issue (short term or powered deployment)

Challenges and opportunities:

or questions not often answered in sensor network research

Challenge 1: Programming

Developing an application on SNs is hard

- Programming is hard for most people anyway
- A WSN is a very complex system (a real system encompasses many layers from the sensor to the back end data processing)

- **In reality two problems:**
  - Programming = no different from other scientific applications (lack of tools for sensor networks and low level languages is an issue)
  - System design and deployment = The data needs to be sent somewhere, stored, indexed, searched, analyzed, mined … this all implies a non-trivial IT infrastructure

Challenge 1a: Actual programming of SN

Much progress made:

- High level languages (TinyDB, Regiment, Kairos, etc.)
- Abstractions and better concepts (Virtual Machines, compilation)
- Supporting infrastructures

… but it is not enough

- Even those solutions require considerable expertise
- Current stack is not reliable enough (no fault hiding, no abstractions)

The key question is who is the customer:

- Scientists?
- End users?
- Computer scientists?

Challenge 1b: Complete IT infrastructure

Related work can be applied

- Data stream processing (Hi-Fi project)
- Data Mining and stream mining
- Data Visualization
- Standard web servers – database techniques

This is not a new problem but nevertheless very important

- Management of scientific data (all aspects)
- Lineage, versioning, archival

The use model needs to be clear:

- On-line, real time monitoring with interactive interface
- Data gathering and simple instrumentation plus archive

Opportunity 1: Turn key solutions

Sensor Networks will be widely used only when

- they stop being just toys for nerds
- the user does not need to understand the technology involved
- the amount of effort required to use a SN is significantly reduced

The best way to accomplish this is through turn key solutions

- develop complete systems from sensor to repository and including the data processing tools
- tailored to one type of applications
- Do the math and engineering: life time, total cost (production, deployment, operation, maintenance, reuse)
- Compare with alternative designs

Challenge 2: Deployment

Sensor networks make sense when there are no easier alternatives:

- Remote locations
- Difficult environments
- Costly installation of alternatives

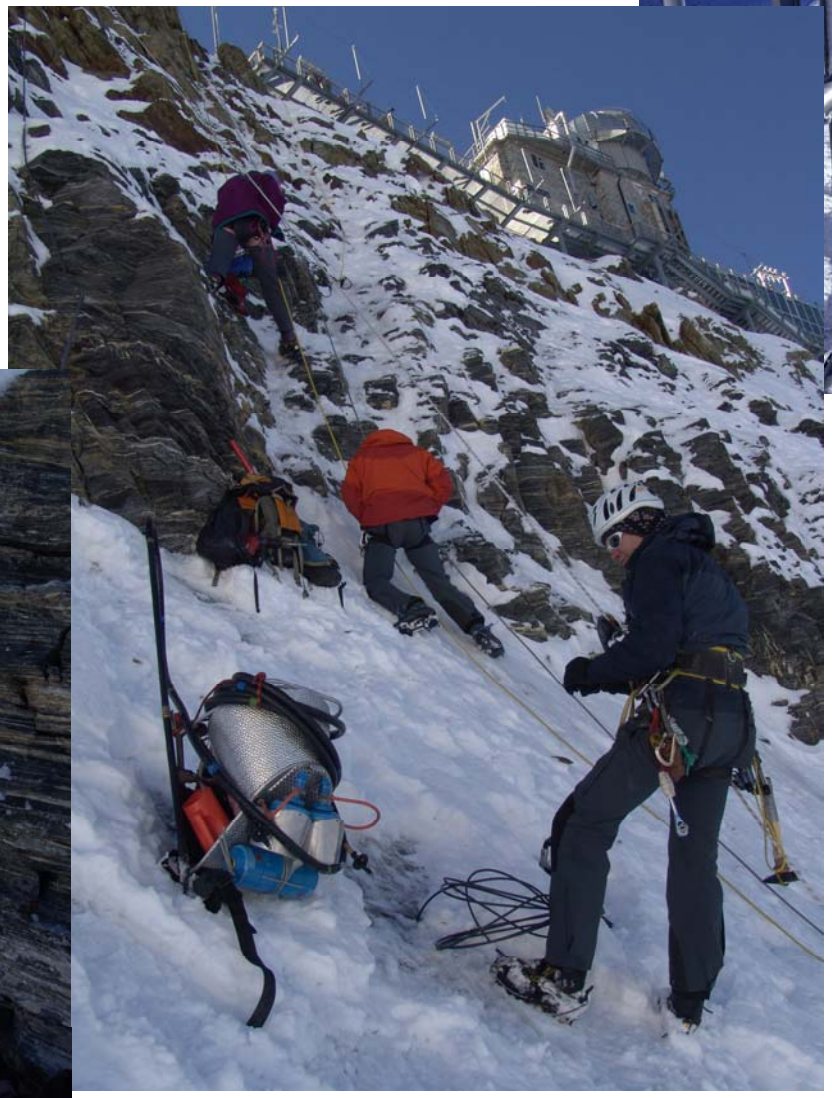
However, the cost of a real deployment can be significant:

"Off-the-shelf WSN solutions currently available in the market are not ready for real deployments for environmental monitoring. They need considerable modifications and improvements in order to meet the requirements of the natural sciences."

PermaSense: Investigating Permafrost with a WSN in the Swiss Alps
4th Workshop on Embedded Networked Sensors (EmNets 2007), Cork, 25-26 June 07

Example: PermaSense

Study the changes in permafrost areas



Every deployment is a custom design

PermaSense:

- Ad-hoc design of the physical sensor platform
- Calibrate the sensors for the necessary accuracy
- Change the TinyNode hardware to reduce power consumption
- Change TinyOS to correctly execute automatic power management
- Add Forward Error Correction to network stack
- Implement caching strategy for reliable delivery
- Implement their own multi-hop routing protocol
- …(for more details: http://cn.cs.unibas.ch/projects/permasense/)

Panel this afternoon: "Sensor Networks, are we done?"

Opportunity 2: Real Deployments

Most of what is out there works only, and if at all, in the lab.

Real deployments show how far we are from being done:

- Data Yield (reliability, correctness)
- Synchronization
- Routing and network dynamics
- Scale
- Packaging and complete solutions
- Calibration
- Programming and maintenance
- Reusability

Challenge 3: In-network data processing

Hot research topic:

- Difficult and challenging
- Algorithmically complex

Not the right issue according to some:

> The Tenet project is developing an alternative architecture for tiered wireless sensor networks which contain both small-form-factor motes and Stargate-class "masters". The Tenet project's guiding architectural principle asserts that multi-node data fusion functionality and complex application logic should be implemented only on the masters, since the cost and complexity of implementing this in motes outweighs the performance benefits of doing so.
>
> http://enl.usc.edu/projects/tenet/

Opportunity 3: Get real specs on systems

The claims that in-network data processing is good or is bad are both based on taste and research interests

- There is no hard data to prove one or the other claim
- The only argument put forward is complexity and reusability
- In some applications, it is really not an issue (all raw data needed)
- When does it make sense:
  - All nodes measuring the same? (only small scale)
  - In deep routing trees? (which are typically dynamic)

We are not making enough of an effort to study the implications of new techniques and ideas we propose

Challenge 4: Sensing the world

The most interesting sensors

- are not the ones in typical sensor platforms
- are expensive
- require expertise to operate
- cannot operate on batteries or only for very short times
- calibration is huge issue

Can it be that wireless sensor networks are limited to sensing:

- Light level
- Noise level
- Temperature
- Voltage

Opportunity 4

Pursue deployments with different constraints

Find what makes sensor networks unique and relevant

- Is it the ad-hoc wireless networking?
- Is it the small form factor?
- Is it the in-network data processing?
- Is it the intelligence on the edge of the network?
- Is it a combination of the above? Which one?

Consider that for all of the above aspects, there are alternative technologies doing the same or better

The road ahead

Large, non-tutored, long term deployments

Data and experience from

deployments as input to research

- Feasible scales
- Real problems
- Life time (MTTF)
- Alternative solutions
- Complete solutions

Aim for real deployments

- Long term
- Reusable
- Cost effective
- Non-tutored
- Well engineered
- Business case

Linking to IT infrastructure

WSNs are today largely ad-hoc:

- Need to connect to IP
- Need to hide peculiarities
- Need standard programming tools
- Need to provide specs
- Need to have a life-cycle
- Need to evaluate complete stack

Heterogeneity and diversity:

- One can argue that future WSNs will not be at all like the ones we are building today
  - WiFi coverage
  - SMS
  - UMTS
  - Mobile devices

Are we solving the problems of Wireless Sensor Networks or of the platforms we have chosen to use?

Conclusions

Computer Science = Science of the Artificial

Sensor Networks = Data from the real world

Working in sensor networks requires a different approach

than the one in conventional computer science

Impact comes from the whole system not from the parts

Questions
(and perhaps answers)