Sensor Data Engineering
Context and Perspectives

Philippe Bonnet, IT U.Copenhagen - phbo@itu.dk
Observation / Monitoring
Sensor Data essentially used to model physical systems. Those models dictate how sensor data should be collected and managed.
Sensor Data Engineering

Engineering techniques and methodologies for the design, development and assessment of data acquisition and sensor data management systems.
Sensor Data Lifecycle

- **Instruments**
  - High-bandwidth – not all captured data part of raw measurements (particle physics, vulcano monitoring)
  - Low-bandwidth – all data are captured (ecological datasets)
- **Raw measurements are primary data, data products are secondary data**
  - e.g., NASA CODMAC
- **Instrument Control Loop**
  - Parameters adjustment based on model knowledge
- **Archiving/Curation of cold data.**
Example Data Products

NASA CODMAC:

**Level-0**: raw engineering measurements (full resolution)

**Level-1**: raw measurements time-referenced and with meta-data computed, appended but not applied

**Level-2**: Derived geophysical variables at the same resolution

**Level-3**: Variables mapped on uniform space-time grid scales

**Level-4**: Model output or results from analyses of lower level data
Outline

Sensor Data Engineering

Sensor Network-Based Instrumentation
Rationale #1

Sampling at unprecedented scale and resolution

Graph courtesy of A. Szalay

Astronomy showing the potential of inexpensive sensor + computing

Graph courtesy of A. Szalay
The MEMS(*) Revolution

(*) Micro Electro-Mechanical Systems

Source: enablingMNT

Evolution in time of the form factor of accelerometers.

Form factor, cost and energy consumption are highly correlated.

Reality Check: Exponential evolution in the size of sensor driven by automotive, consumer goods and medical devices industries. True for Accelerometers, gyroscopes, imagers, microphones. Not true for optical, Chemical, biological sensors.
Rationale #2

Sensornet-Based Instrumentation (a) introduces automatic observations in remote places where only manual observation have been possible so far, and (b) transforms stand-alone devices into a networked system that is monitored and controlled for better performance.
Mote-Based Instrumentation

System challenges: (1) Prototyping tomorrow's platforms
(2) Exploring the boundaries of today's platforms
Design Space

1. Choose HW components
2. Choose SW components (incl. OS)
3. Implement new components
Design Space

Data Management

Sensor network regime
- Connectivity vs. Duty cycling
  - Low-Power Listening
- Topology control
  - Gateway + cluster of 20-50 motes
- Network abstraction
  - IP (6LowPan)
- Routing: open issue
Software Development (Data Acq.)

- Programming abstractions
  - **Concurrency**: thread vs. events, time representation
  - **Distribution**: primitives (reliable dissemination, collection tree), macroprogramming

- Testing: controllability and observability are hard/expensive to enforce when instrumenting a physical system. Verifying calibration, testing compliance after deployment are open issues.

- Health management: making sure data acquisition functions and performs as required once deployed is an open issue.

- Design Principles: Equivalent of Rules of thumbs in data engineering [Gray], Hints of system design [Lampson]
Data Acq./Mgt Paradigms

Sensor specific

- Prob. DB [MayBMS, Trio]
- StreamDB [Aurora, Stream, ...]
- Autonomous Data Acquisition [iDare]
- model-based query processing [BBQ, COLR]

separated

- Gateway-based Systems [SwissQM, Tenet]
- RDBMS

integrated

- Delay-tolerant query processing [IceDB]
- In-network query processing [Cougar, TAG]

generic
Software Development (Data Mgt.)

- Interesting times!
  - Tape is dead, disk is tape, flash is disk [Gray]
  - RAM is disk, L2 is RAM on multicore [Ailamaki]
  - Drop disks and forget about buffer manager, locking and logging [Stonebraker]

- Sensor data are uncertain spatial time series
  - Impacts data model, query language [Widom]
  - Impacts storage structures [Madden]

- Loading from Data Acq to Data Mgt is key [Howe]

- Supporting the complete sensor data lifecycle is an open issue.