A Semantic-based Architecture for Sensor Data Fusion

UBICOMM ‘08

Stamatios Arkoulis, PhD Candidate

National Technical University of Athens

Outline

- Introduction
- Existing Systems & Applications
- Data Transformation & Semantic Representation
- Motivation
- The Proposed Architecture
  - Data Layer
  - Processing Layer
  - Semantic Layer
- Conclusions
- Future Work

Introduction

- Sensor networks (wireless)
  - attract a lot of attention → research → innovation
- Many implementations worldwide (why?)
  - Nature: smallness, price, energy-efficiency, reliability
  - IPv6: no number limitations
    - huge address space for networking purposes
  - Improvements in resource management (battery life, computation & communication capabilities, memory)
- What about the future?
  - sensors everywhere
  - accessible via the Internet

Introduction

- Main scientific attention
  - networking of distributed sensing
  - But what about…
    - management, analysis, understanding of collected data
- Current sensor deployments
  - Huge number of sensors
  - Heterogeneity
  - Rawness of data
    - little, if any, meaning by themselves
  - Enormous amounts of data stored

We have *many* "Data" but *little* "Information"

Introduction

The Solution ?

- Proper data management (for event detection)
  - Data interpretation
    - considering final user requirements / needs / scope
  - Data heterogeneity
    - special aggregation schemas
    - combination / comparison / correlation
  - Data Aggregation & Processing
    - render them helpful to applications

Key Technology: “The Semantic Web”

Introduction

- The Semantic Web
  - despite the heterogeneity and amount of collected data
    - "*meaningful*" events extraction
    - interoperability

"***<u>Connection of Sensory Data to their environmental features</u>***"

- "*Semantic Annotation*" → What?
  - metadata added to any form of content
  - well-defined semantics ease its use
- "*Semantic Annotation*" → How?
  - content description languages
  - query languages
  - annotation frameworks

Existing Systems & Applications

- Sensor Web Enablement(SWE) initiative [by OGC]
  - Goal → Development of Standards
    - discovery/exchange/processing of sensor observations
  - Common encoding/transport protocol used by services
    - no explicit ontological structure proposed yet
    - no formal conceptual model → no interoperability
- ES3N architecture
    - Semantic Web technologies on top of sensor networks
    - sensor observations' ontology-based storage
      - mechanisms – RDF Repository
    - end-user posts semantic queries (SPARQL)
    - scalability – large data volumes
      - questionable performance / efficiency

Existing Systems & Applications

- IrisNet Architecture – Software infrastructure
  - data collection/storage organization (XML) → Agents
  - end-user queries vast quantities of data (XPath)
  - Scalable – Powerful – Efficient
    - million/distributed/high bit-rate/heterogeneous sensors
  - No semantics → No data reusability
- SWAP 3-tier architecture (framework)
  - sensory data combination (for high-level tasks)
    - unified end-user view of underlying sensor network
  - Sensor / Knowledge / Decision layer → Agent-based
  - services semantically description to end-users
  - multiple agents and ontologies → Complexity

Existing Systems & Applications

- Priamos Middleware – Architecture
  - automated
  - real time
  - unsupervised annotation of low-level context features
  - mapping to high-level semantics
  - rules composition through specific interfaces
    - content annotation without need of technical expertise
    - context awareness challenges easily addressed

Data Transformation & Semantic Representation

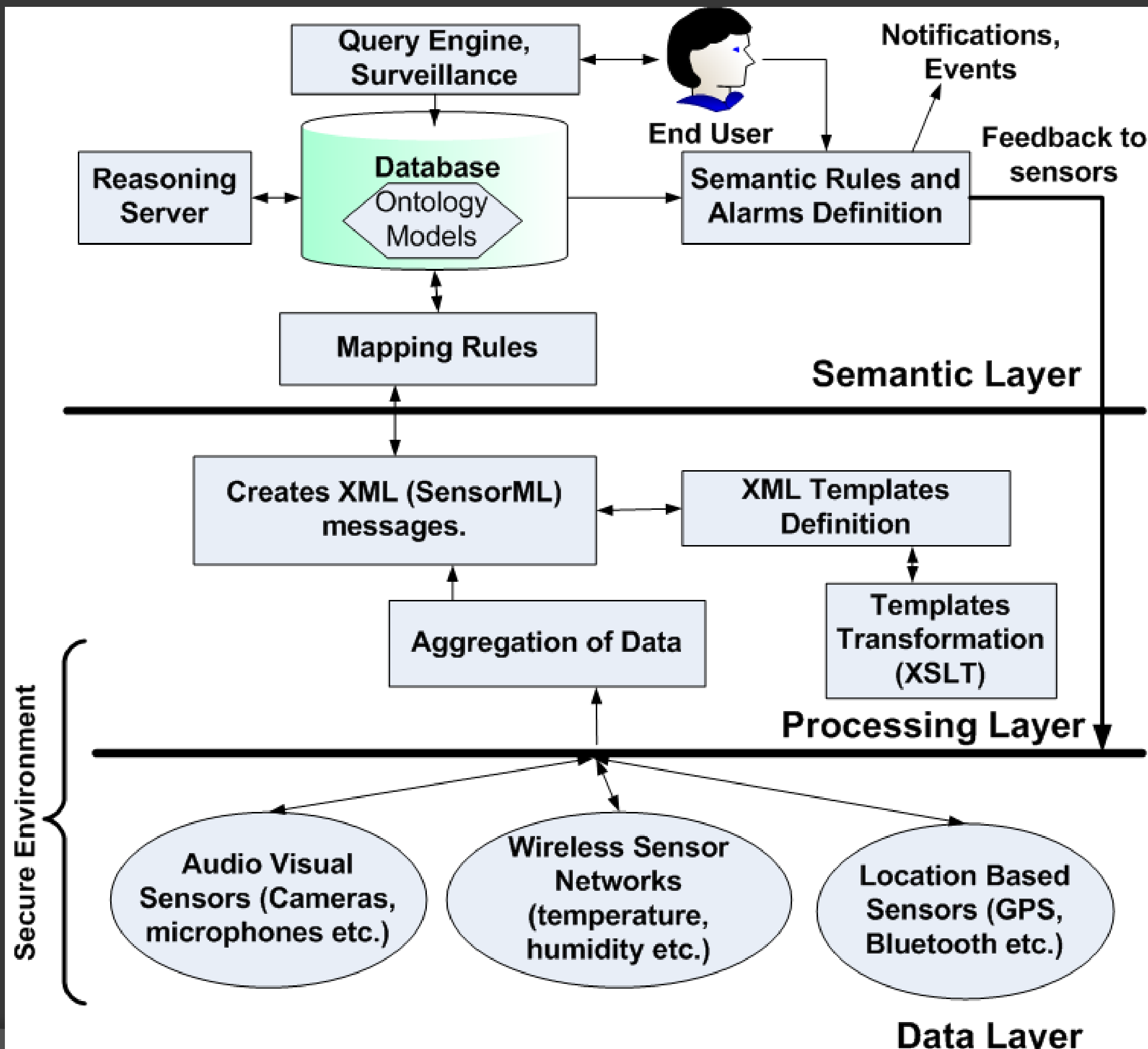
- Specific interpretation models
  - makes raw sensory data meaningful
  - apps' scope / kind of information dependent
- OGC specifications → Models & XML Schema:
  - *O&M*: sensor observations and measurements (both archived or/and real-time) encoding
  - *SensorML*: sensor systems (i.e. location) and sensor observations' associated processing description
  - Formal semantic representation offers…
    - structured knowledge (concerning a certain domain)
    - specify a domain's important concepts/relations

Motivation

- Existing Sensory Data Management approaches
  - no support for distributive sensor deployments
  - inability to scale well in today increasing environment
    - no applicable to large sensors network
    - small amount of data can be transferred
  - lack of context annotation / semantic data representation
  - absence of ontological infrastructures for rules & queries
- Such limitations obstruct end-users to…
  - fully exploit the acquired information
  - match events from different sources
  - deploy smart apps able to follow semantic-oriented rules

“***Solution: A completely new architecture***”

The Proposed Architecture



- Data management
  - Gathering
    - Real-time
    - Recorded
  - Aggregation
    - Heterogeneity
  - Processing
    - Meaningfulness
- User defined rules
  - alarms - actions
- Flexibility
- Modularity
- Scalability

Data Layer

- Central entities
  - sensor discovery
  - data acquisition (+policies)
    - event-based → data sent directly
    - polling-based → data periodical queried
  - raw-data collection
    - Location Sensors (Smart Phones, PDAs, etc.)
      - positioning interfaces (Bluetooth, GPS, Wi-Fi)
      - location-sensitive data
      - next layer reached either directly or via special infrastructures
    - Wireless Sensors

Data Layer

- Wireless Sensors (MICA2, eKo, Imote2)
  - Measure and monitor environmental metrics
  - Architectures, routing protocols and schemes exist
    - efficient energy consumption & congestion avoidance
  - Data reaches the next layer
    - routed to specific nodes & forwarded to central entities
    - send directly to central entities
- Audiovisual Sensors (microphones, cameras, etc.)
  - rich, real-time content
  - special networking requirements to be satisfied
    - bandwidth (huge amount of bits to be transmitted)
    - packet loss (destroyed content / wrong order)
    - jitter (glitches)
- Data reaches the upper layer → Web services etc.

Data Layer – Security Issues

- Security Requirements
  - Data confidentiality/integrity/freshness/authentication
  - Secure time synchronization / localization
  - Anonymity (hide location of sensor-observed aspects)
  - Secure transmission between Sensors-Aggregators
    - Secure Web Services, SSL, X.509, PKIs, XML encryption
- Obstacles to Security
  - resource / computing constraints
  - communication reliability
  - unattended operation
- *Optimality*: Safety – Efficiency “trade-off”
  - Sensors’ type & Deployment scenario dependence

Processing Layer

- Aggregators (due to sensors' limited resources)
  - raw data processing
  - data transformation to useful ("standard") formats
    - XML generation
    - dynamic system configuration through XML schemas
      - sensors' capabilities/location/interfaces formal descriptions
      - specification of different data significance for users' apps
    - XML files re-transformation (XSLT Module)
  - XML files forwarding to the upper layer
- GSN (Open Source – Java)
  - user-defined wrappers (based in a data model)
  - incoming data encapsulation to the data model

Semantic Layer

- Abstraction of received “XMLs”
- Context capturing in varying conditions
- “Automatically” configured context annotation
  - by application specific ontologies

- This layer consists of
  - an exported Web Service interface
  - ontology Models
  - Mapping and Semantic Rules
    - …and the corresponding actions / notifications
  - the external Reasoning Server

Semantic Layer

- Web Service interfacing module
  - messages (from the lower layer) manipulation
    - any arbitrary well-formed XML document
    - knowledge is transferred
- Ontology models
  - Database Model → Jena internal graph engine
  - Ontological Model → Triple statements
  - Knowledge Base → Annotation (separate from data)
  - Incoming XML files stored
    - transformation in another XML template

Semantic Layer

- Rules (syntactic and semantic homogeneity)
  - Knowledge conversion into semantic information → KB
  - XML Mapping Rules
    - fetch data from XML message
    - storing in ontology model as ontology class individuals
  - Semantic Rules
    - modify the ontology model
  - Distinction inspired by RuleML
    - RDF-only and RDF-XML-combining subsets
    - common syntax
      - different conditions & actions in each case
  - Event-Condition-Action pattern followed
    - **“on** *event* **if** *condition* **then** *action”*

Semantic Layer

- Mapping Rule
  - IF EXISTS /sensor/temperature/@value THEN INSERT INDIVIDUAL IN CLASS Temperature AND SET DATATYPE PROPERTY hasValue /sensor/temperature/@value
- Consecutive Semantic Rule
  - IF DATATYPE PROPERTY IN CLASS Temperature HAS VALUE GREATER THAN 40 AND DATATYPE PROPERTY IN CLASS Humidity HAS VALUE GREATER THAN 0.3 THEN Alert (”Surveillance area under unusual conditions!”)
  - Trigger Alerts based on KB awareness of the world
  - Semantic-based intelligence added
    - reasoning procedures deduce implicit knowledge based on the current explicit facts

Semantic Layer

- Reasoning server
  - Knowledge Base is Ontology-Reasoner combination
    - Reasoner (essential)
      - OntoBroker, KAON2, Pellet etc.
      - DIG interoperability / Stand alone DIG servers
        - HTTP message exchanging with calling programs
      - Jena *supports* biding of external reasoners

"***choice is up to the user***"

Conclusions

- Modular architecture for deploying WSNs
  - ease end-user to take advantage of collected data
  - facilitate developers
    - deploy new useful applications
    - exploit the Semantic Web advances
  - add flexibility to the sensor world
  - form associations over the raw data
  - extract meaningful information and valuable results
  - create specific management & notification rules
    - based on the nature of applications

Future Work

- Implementation of different scenarios
  - combine aggregation/security/processing methods
- Evaluation of architecture’s discrete components
  - Scalability & Performance issues
- Study energy efficiency trade-offs under
  - proposed routing schemes
  - data aggregation architectures

Questions?

Thank you for your attention !!

Stamatios Arkoulis

stark@cn.ntua.gr