Novel System Architectures for Semantic Based Sensor Networks Integraion

ZORAN BABOVIC, ZBABOVIC@ETF.RS
VELJKO MILUTNOVIC, VM@ETF.RS
THE SCHOOL OF
THE ELECTRICAL ENGINEERING,
UNIVERSITY OF BELGRADE

Introduction

- The progress in the sensing and wireless technology implies proliferation and deployment of various sensor nodes devices
- Sensor Web – the vision of heterogeneous sensor networks integration on the Web
- Semantic Based Integration – Ability of independent systems to exchange, understand, interpret, and process data produced by other systems based on semantic data
- Semantic Sensor Web is a platform that enables possibility for providing more complex services by supplying context-related information with the raw sensor data

Challenges of Sensor Networks Integration

- Basic organization
- Scalability
- Sensor data sources heterogenity
- Flexibility of supported sensor networks
- SN capability awareness
- SN management and actuation functions
- Ontologies and level of applied semantics
- The data representation model
- Query language
- Knowledge inference
- Application/Service interface and data format
- Service discovery
- Service composition
- Quality of service and information
- Security

Clasiffication Criteria for Existing Approaches

- Two main available approaches:
  - Bottom-up – Sensor Networks Oriented Approaches
  - Top-down - Application Oriented Approaches
- Sensor Networks Oriented Approaches include:
  - Database Centered
  - The Query Translation
  - Sensor Virtualization
- Application Oriented Approaches include:
  - The Service-Oriented Architecture
  - Service-Composition Oriented Approach
  - The Rule Based Data Transformation
  - The Agent Based Systems

Database Centered Solutions

- Basic characteristics:
  - A database is a central hub of all collected sensor data
  - All search and manipulation of sensor data are performed over the database
- Challenges:
  - How to map heterogeneous sensor data to database schema
  - Support for real-time data provision
  - Scalability
- Advances:
  - The ability for applying data mining algorithms over stored data in order to extract additional knowledge
- Non-Semantic Solutions
  - Cougar, Cornel University in 2001
  - SenseWeb, Microsoft Research, 2008
- Semantic Based Solutions
  - E3SN, University of Georgia, 2006

Query Translation Approaches

- Basic characteristics:
  - Users query are transformed to the target query language of certain sensor data source
  - The results of native queries should be assembled into the target data format
- Challenges:
  - Maintenance of information of available data sources, primarily the native query language of certain data source, format and nature of produced data, but it may also include information about sensors capabilities, network topology, power constrains for better query optimization
  - Potential performance drawback – two data conversion per one request
- Only Semantic Solutions:
  - CSIRO SSN, by CSIRO ICT Centre, Australia in 2008
  - $\text{SPARQL}_{\text{STREAM}}$, Polytechnic University at Madrid and University of Manchester, 2010
  - SemSorGrid4Env, EU FP7, 2011

Sensor Virtualization Approaches

- Basic Characteristics
  - Sensors and other devices are virtualized with an abstract data model
  - Applications are provided with the ability to directly interact with such abstractions using the specified interface
  - Multiple levels of sensor data formats might coexist depending on the user needs
- Challenges:
  - Produced data streams must comply with the commonly accepted format that should enable interoperability
- Non-Semantic Solutions
  - GSN (Global Sensor Network), EPFL, Switzerland in 2006
- Semantic Solutions
  - SENSEI, EU FP7, 2010
  - IoT (Internet of Things), EU FP7, 2012

Service-Oriented Architectures

- Basic Characteristics
  - Provide standard service interface with defined methods and data encodings for obtaining sensor observations and measurements
  - There could be offered functions for subscription on sensor data, performing actuation functions and others
  - Dominant interaction model is request-reply, and to a lesser extent the event-based delivery of sensor data
- Challenges
  - How to fuse stream-based sensor data with aquisitional and archived sensor data
- Non-Semantic Solutions
  - TinyREST, by Fraunhofer and Samsung in 2005
  - SWE, by OGC (Open Geospatial Consortium) in 2006
- Semantic Solutions
  - SemSOS, by Wright State University in 2009

Service-Composition Approaches

- Basic Characteristics
  - Offer to users the ability to define arbitrary services or data streams with specific characteristic they are interested in
  - The system will try to compose such a data flow by applying specific processing over appropriate data sources, which will result in producing a data stream that conforms to the requested specification
- Challenges:
  - How to describe sensor data sources and processing elements in order to enable efficient reasoning and composing of desired data streams
- Non-Semantic Solutions
  - Hourglass by Harvard University in 2004
- Semantic Solutions
  - SONGS by Microsoft in 2005
  - System S Middleware by IBM in 2007

Rule Based Data Transformation Approaches

- Basic Characteristics:
  - Mapping functions are based on the relationships between the concepts captured in the ontological representation of the domain model and sensor data observations and measurements
  - Data are transformed from lower level formats to semantic-based representations that enable semantic search over available data and applying of reasoning algorithms
- Challenges:
  - An appropriate information model should be designed in order to cover various application domains and sensor devices
- Semantic Based Solutions
  - Semantic Based Data Fusion by University of Toronto in 2007
  - Data Transformation by Mapping Rules by National Technical University of Athens in 2008
  - SWASN by Ericsson in 2008

Agent Based Systems

- Basic Characteristics:
  - There are several types of agents,
    software components capable of performing specific tasks,
    which collaboratively achieve desired functionalities
  - Typically, agents belong to one of several layers depending on the type of functionalities they are responsible for
  - Agents from upper layers employ agents from lower layers
- Challenges
  - How to model agent processing capabilities and
    internal data formats
- Non-Semantic Solutions
  - IrisNet by Intel and Carnegie Mellon University in 2003
  - SWAP, by ICT for Earth Observation Research Group in 2006

Comparison

- **Scalability:** The most scalable approach is directory based approach used in sensor virtualization architectures
- **Users' flexibility**: Most comfortable approaches from users' perspective are service-composition architectures
- **Information model**: The most comprehensive information model has been proposed in large EU FP7 Projects Sensei, IoT, and SemSor4Grid based on the W3C's SSN Ontology
- **Application interface**: The REST interface is the most efficient application interface implemented in many solutions

Authors' research efforts

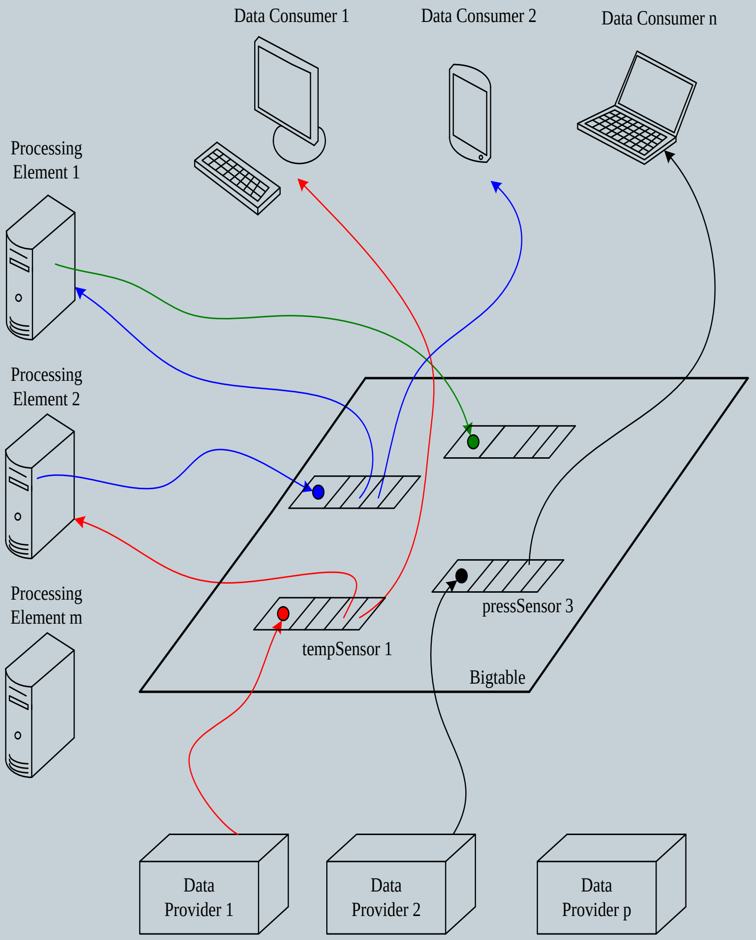
- Traditional RDBMS fail in supporting the management of high volume of sensor data provided by multiple data providers and huge number of Internet users
- NoSQL database systems offer high data availability while maintaining petabytes of data distributed over thousands of commodity machines
- Column stores are widely used for Internet scale applications by Google, Facebook, Twitter and others
- We have considered a column store distributed repository for keeping obtained sensor data represented through RDF, and performing search over such data using appropriate indices
- This platform can be used for publishing Linked Sensor Data, following principles of Linked Data
- Our prototype is based on the HBase column store built on the top of the Hadoop

Mapping of RDF sensor data to column store

- In column stores, there is no strict schema for columns which can be dynamically added or removed
- Sensor data are represented as triples <*s*, *p*, *o*> using RDF
- How to support all triple search patterns ?
- Subject Centered Indexed Table: one subject per row with multiple predicates mapped as columns
- Predicate Index Table: <*po_s*> type of index, where *predicate_object* is a row key, and a subject as column key
- Separate table for keeping spatial data index
- Temporal information of sensor observations are coded as URI concatenation with the timestamp

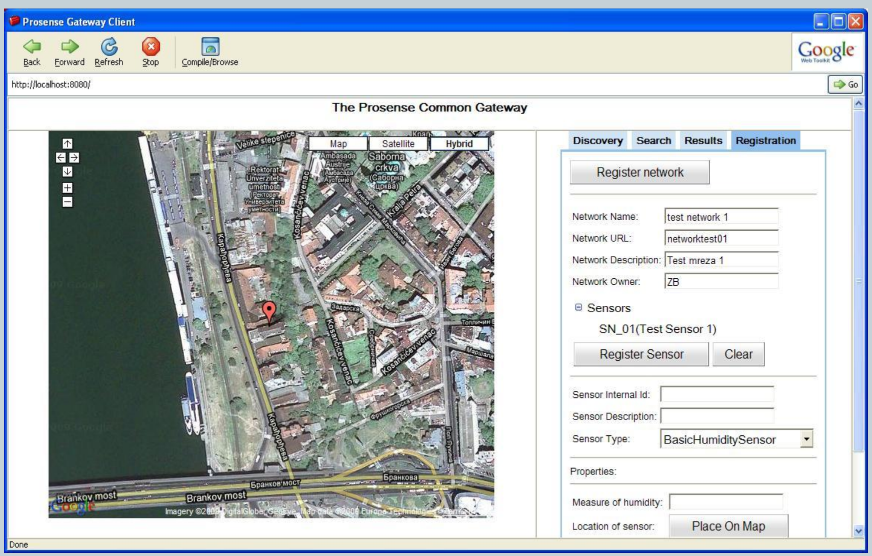
The Architecture based on the column store

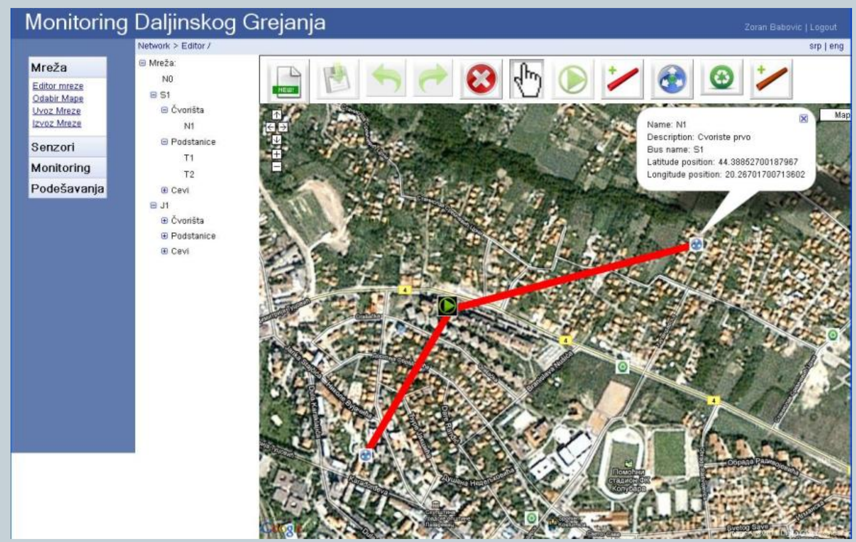
- Ability of applying distributed multiple processing on sensor data using MapReduce
- Users are able to subscribe on either the data from the sensor data source or to complex data streams published by processing elements
- Spatial-temporal search is improved using separate index structures



Sensor Web Applications

- Public District Heating Monitoring System
- Semantic sensor data search





Future research directions

- Publications of sensor data as linked resources using XLINK mechanism
- Some researchers investigate extensions of available semantic query languages
- Creation of a flexible information model that will satisfy needs of many sensor application scenarios
- Investigation of efficient distributed structures suitable for managing spatial-temporal characteristic of huge sensor data volume
- Data mining and processing of Big Sensor Data