Real Time IoT Stream Processing and Large-scale Data Analytics for Smart City Applications

Ralf To¨njes∗, Muhammad Intizar Ali†, Payam Barnaghi‡, Sorin Ganea¶, Frieder Ganz‡, Manfred Haushwirth†, Brigitte Kjærgaard‡‡, Daniel Ku¨mper∗, Alessandra Mileo†, Septimiu Nechifor††, Amit Sheth‖, Vlasios Tsiatsis∗∗ and Lasse Vestergaard§

∗University of Applied Sciences Osnabru¨ck, Germany, †Insight @ National University of Ireland, Galway, ‡‡ City of Aarhus, Denmark, ‡ University of Surrey, Guildford, UK, †† Siemens Romania, §Alexandra Institute, Aarhus, Denmark, ¶Brasov Metropolitan Agency, Romania, ‖Wright State University, Ohio, ∗∗Ericsson Research, Sweden

Motivation and Goal

- The achievement of sustainability for smart city frameworks is challenged
  - integration across different application domains
    - valuable data and sensory information are limited to specific application domains due to specific formats
    - aggregation of information is typically done manually and is often outdated or is provided static sources of data
  - integrate Internet of Things and the Internet of People
- Reliable real-time smart city applications can be facilitated by knowledge-based computing and reliability testing

Sensory data streams

- Smart city data can be seen as big data; however it is not only large in volume
  - rapid changes in data and dynamically of the environment
  - resource constraints in sensing platforms and distribution and heterogeneity of data
  - trust and reputation that need to be identified and associated to these sources
  - efficient stream reasoning mechanisms are required to interpret the meaning of events in a context-aware fashion, and share such meaning across applications

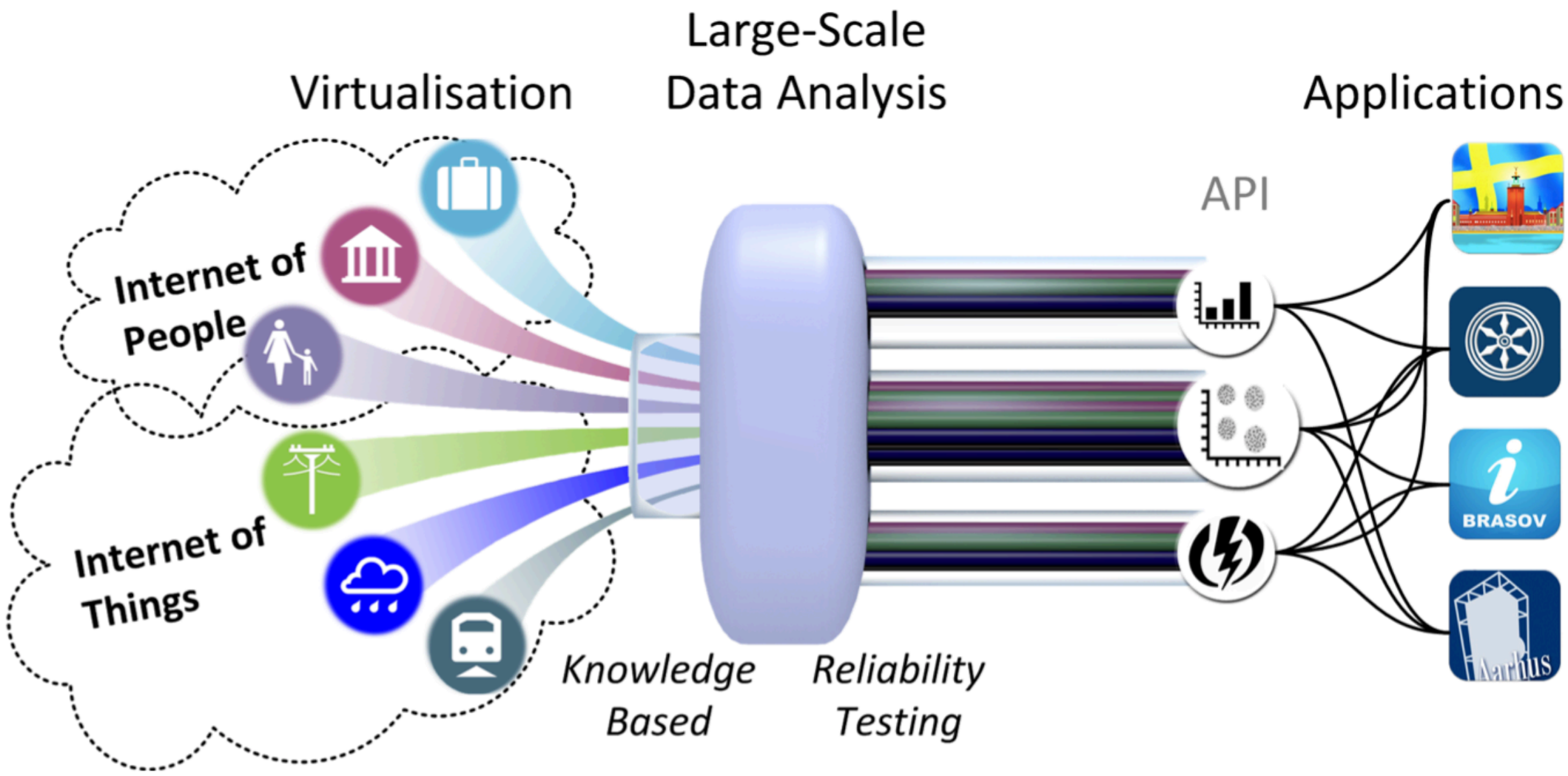

Fig. 1. Integrated approach for IoT and social media stream based smart city applications

Federation of heterogeneous data streams using semantic description and annotation

- Effective utilization by integrating various IoT streams with the existing data sources
  - seamlessly integrate real world data streams
  - automated search, discovery and federation of data streams
  - adaptive techniques to handle fail-overs at run-time

- Semantic Sensor Network
  - focus on IoT resources and data but less on IoT streams and their features

Large scale IoT processing and data analytics

- Efficiently process large scale IoT streams
- Need efficient methods to perform data analytics in dynamic environment by aggregating, summarizing and abstracting sensor data on demand

- Current analytic frameworks have to be evaluated for applicability in the smart city environment and the impact on privacy has to be taken into account

Real-time IoT information extraction, event detection and stream reasoning

- Extraction of high level knowledge form heterogeneous, multi-modal data streams
- Current techniques of stream reasoning do not cater to the needs of IoT due to the lack of proper treatment of uncertainty
- Existing event detection and processing techniques are typically limited to a single one modality
- Lack of an event detection approach that would take into account multiple data streams considering their varying nature in terms of trustworthiness, reliability

Reliable information processing, QoI, testing and monitoring

- To evaluate accuracy, trustworthiness, and provenance of IoT streams
- To resolve conflicts in case of contradictory information
- Continuous monitoring and testing to dynamically update QoI and trustworthiness

- New challenges of QoI due to distributed ownership, high dynamically and heterogeneity of the systems involved
- Current testing approaches in smart city environments are not able to rely on collected test data

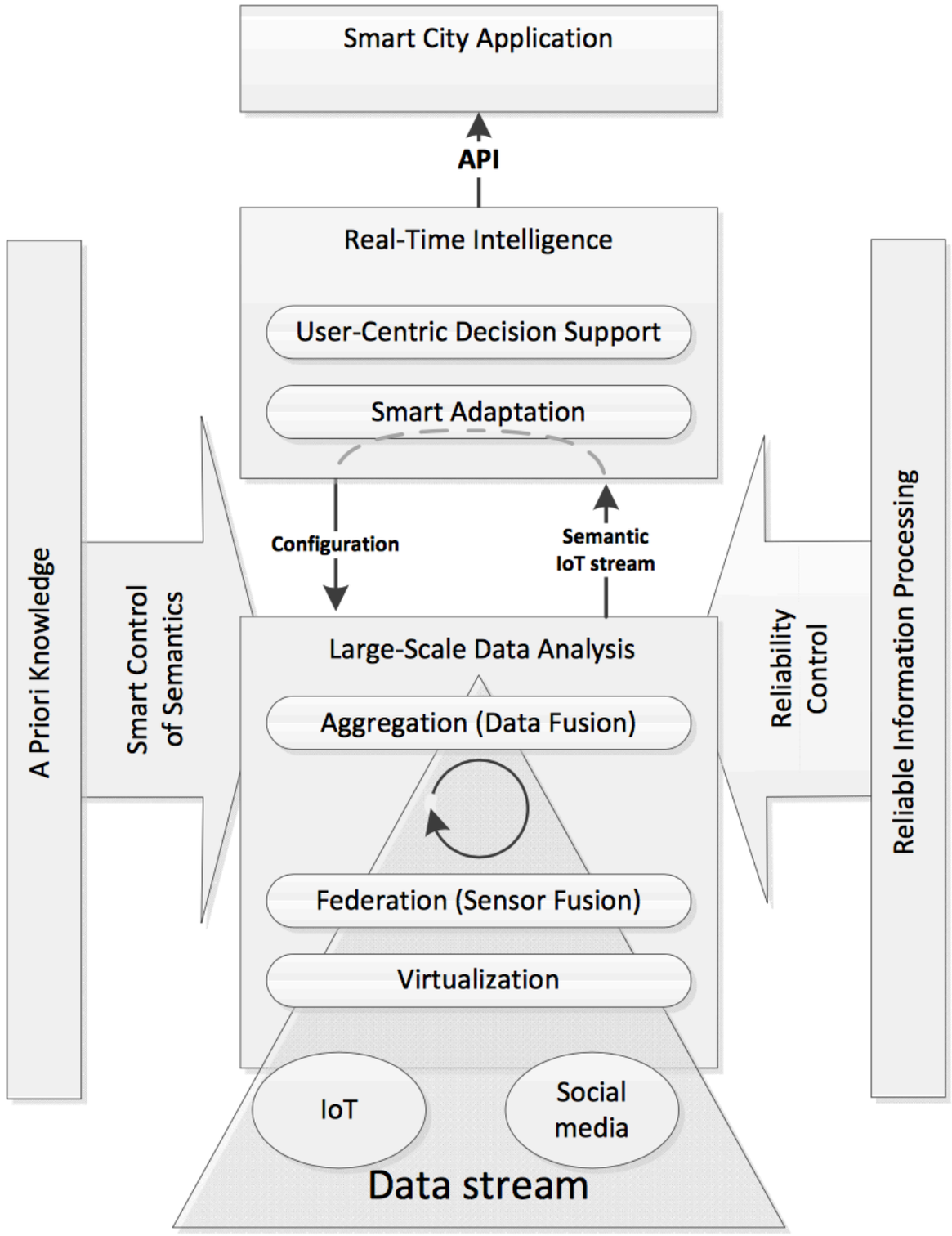

Fig. 2. Framework for data stream based smart city applications

CityPulse framework

- Virtualization
  - Semantic annotation of heterogeneous data for automated discovery and knowledge-based processing
- Federation
  - On demand integration of heterogeneous Cyber-Physical-Social sources
- Aggregation
  - Large-scale data analytics

CityPulse framework

- Smart Adaptation
  - Real-Time interpretation and data analytics control
- User centric decision support
  - Context aware customized IoT information extraction
- Reliable Information Processing
  - Testing and monitoring accuracy and trust
- Smart City Applications
  - Application programming interface for rapid prototyping

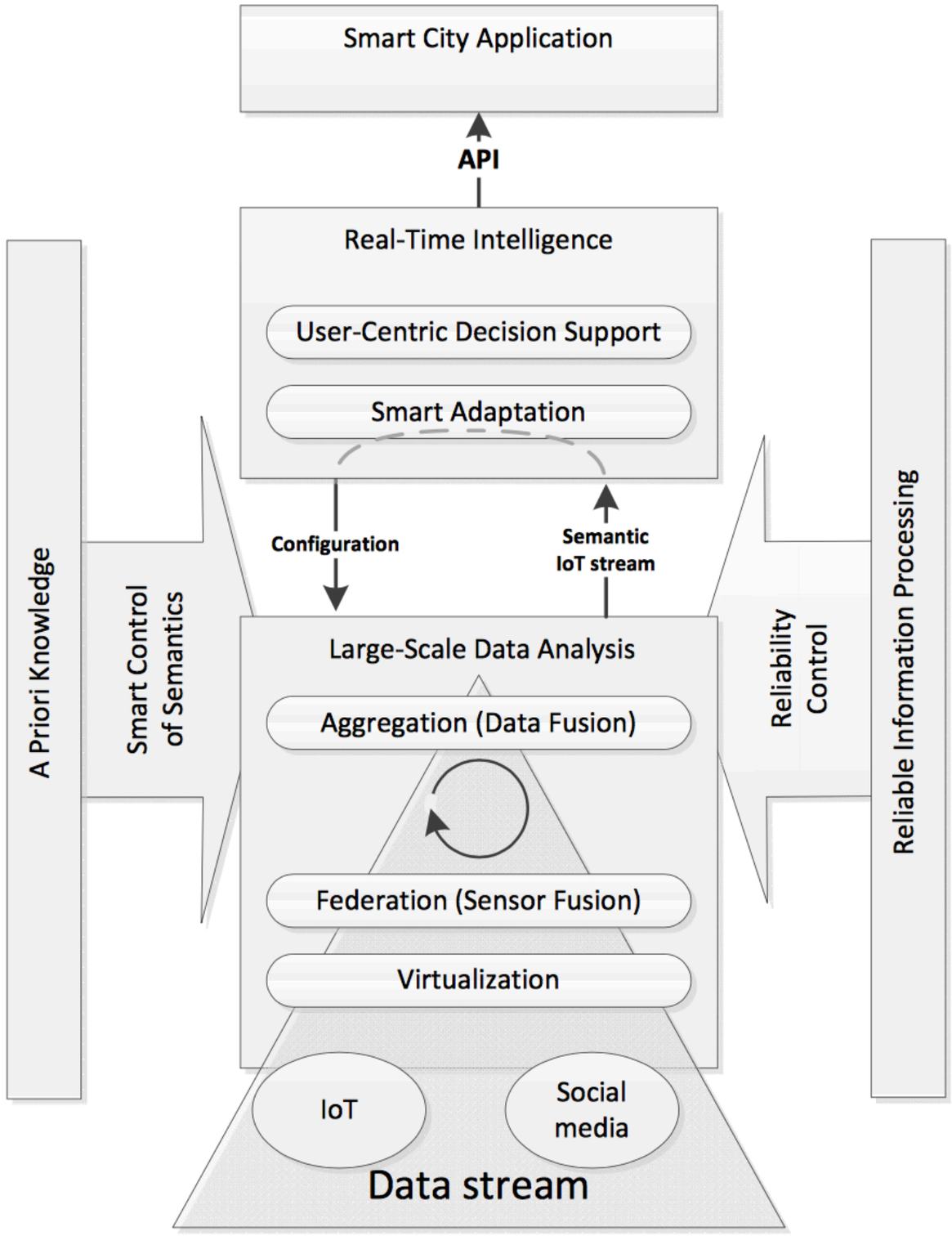

Fig. 2. Framework for data stream based smart city applications

Large-Scale IoT Stream Processing

- Provides an integrated platform to virtualize resources and provide uniform access and integration support irrespective of the information source
- Using semantic annotations for all data and data streams, data becomes machine understandable and processable
- This meta-knowledge is employed to automate the discovery and processing control for data analytics, facilitating event detection, aggregation and reasoning

Reliable Information Processing

- QoI Datastore and Reputation Systems
  - automated rating of information against accuracy, trustworthiness and QoI
  - not only require initial training phase but also require to be adjusted in the later phases by the continuous analysis of data streams
- Testing and Monitoring Environment
  - proactively maintain conflict resolution and fault recovery

Real-Time IoT Intelligence

- Adaptivity
  - real-time adaptivity will be based on the identification of unexpected events
  - suggestions will be made for the dynamic reconfiguration of stream discovery and aggregation mechanisms
- User-centric stream reasoning
  - provide optimal configuration of smart city applications
  - users can specify their requirements and preferences explicitly
  - data mining techniques also play a role in building user preference models resulting in an individualized experience

Traffic management as a support element for sustainable development of the urban fabric

- Brasov (capital of Romania)
- Integrate knowledge about GPS location and video surveillance
  - traffic conditions
  - organize and optimize traffic flows
  - public transport e-ticketing system
  - management of traffic lights
  - quality of air in the most congested areas of the city

➡ provide users a personalized support to decisions related to transportation.

Open innovation platform

- Aarhus (second largest city in Denmark)
- Give anyone free access to play with public city data, fostering innovation and creation of new services for citizens
  - public trash bin weighings
  - flow of library rentals
- Ability to aggregate and process them to obtain higher-level insights and dynamic monitoring of certain trends are extremely valuable , since it allows to analyze how data is used, by whom and for what

Conclusion

- They discuss principles of large-scale data analytics for real-time smart city data processing and interpretation
- And discuss how various sources of raw sensory data can be combined and processed to extract actionable-knowledge