

# Theta Architecture: Preserving the Quality of Analytics in Data-Driven Systems



Vasileios Theodorou

Staff R&D Engineer, Intracom Telecom

[theovas@intracom-telecom.com](mailto:theovas@intracom-telecom.com)

# Outline

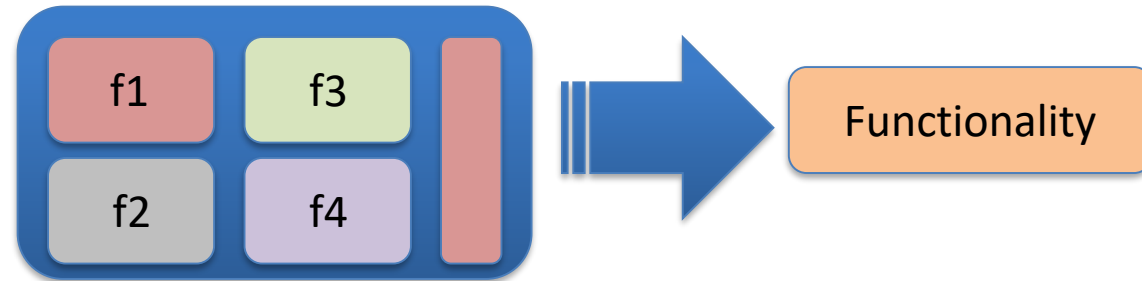
- **Data-driven Systems**
- **Traffic Management Use Case**
- **Theta Architecture**
- **Theta in Action**
- **Discussion**

# Outline

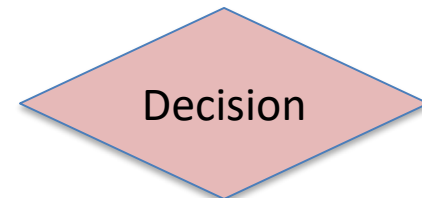
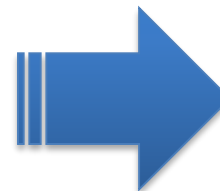
- **Data-driven Systems**
- Traffic Management Use Case
- Theta Architecture
- Theta in Action
- Discussion

# Data-centricity

- From this...



- To this...





# Decisions based on Data Availability

- **Traditional systems**
  - Software-centric
  - Functionality-based
- **Data-driven systems**
  - Decisions based on availability of data
  - Offline and real-time analysis
  - Performance is a must, but not everything
  - Quality of decisions = **F**(Quality of input Information)

# Challenges

- **Big data volume**
  - Tackled with tera/peta-bytes of data processed fast
  - Noteworthy achievements already
- **Other “Vs”**
  - Data scarcity
  - Data quality
  - Low veracity of available data
  - Privacy constraints
  - Tackle all these at runtime

# Idea

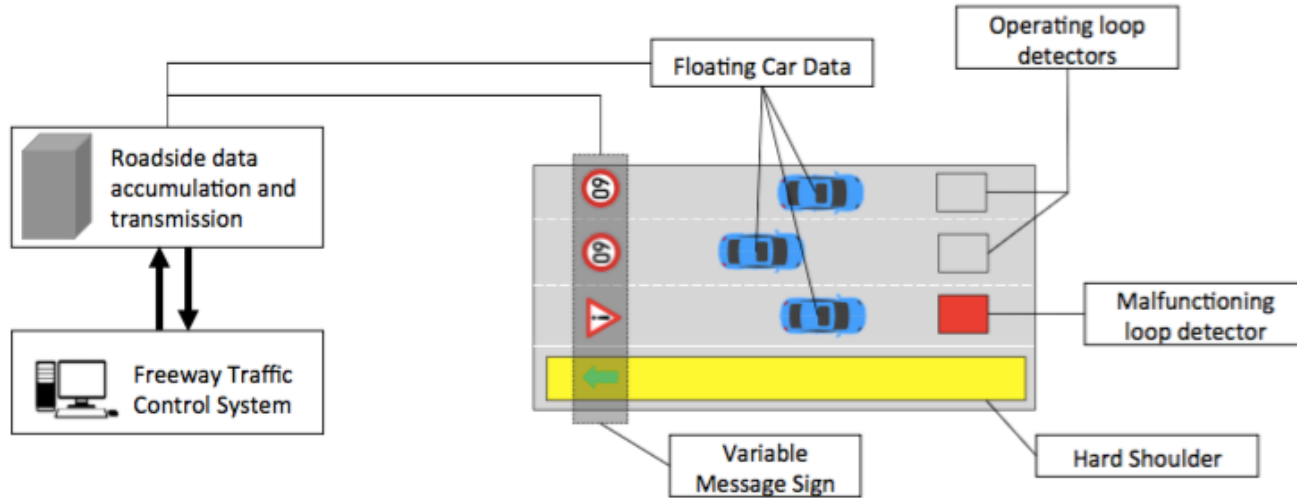
- **Introduce system-level adaptiveness**
- **Decouple business adaptiveness from analytics adaptiveness**
  - Analytics as a first-class citizen
- **Ability to follow data evolution**
  
- **Roadmap:**
  - Identify requirements via examples 
  - Propose high level architecture 
  - Evaluate architecture (future work)

# Outline

- Data-driven Systems
- **Traffic Management Use Case**
- Theta Architecture
- Theta in Action
- Discussion



# Freeway Control System



- **Goal: “Increase freeway capacity”**
- **Means:**
  - Change speed limit
  - Open/close lanes
  - In-car navigation
- **When and how:**
  - Based on data from different sources
    - Loop detectors, cameras, cars’ data, radars, sensing devices, environmental detectors

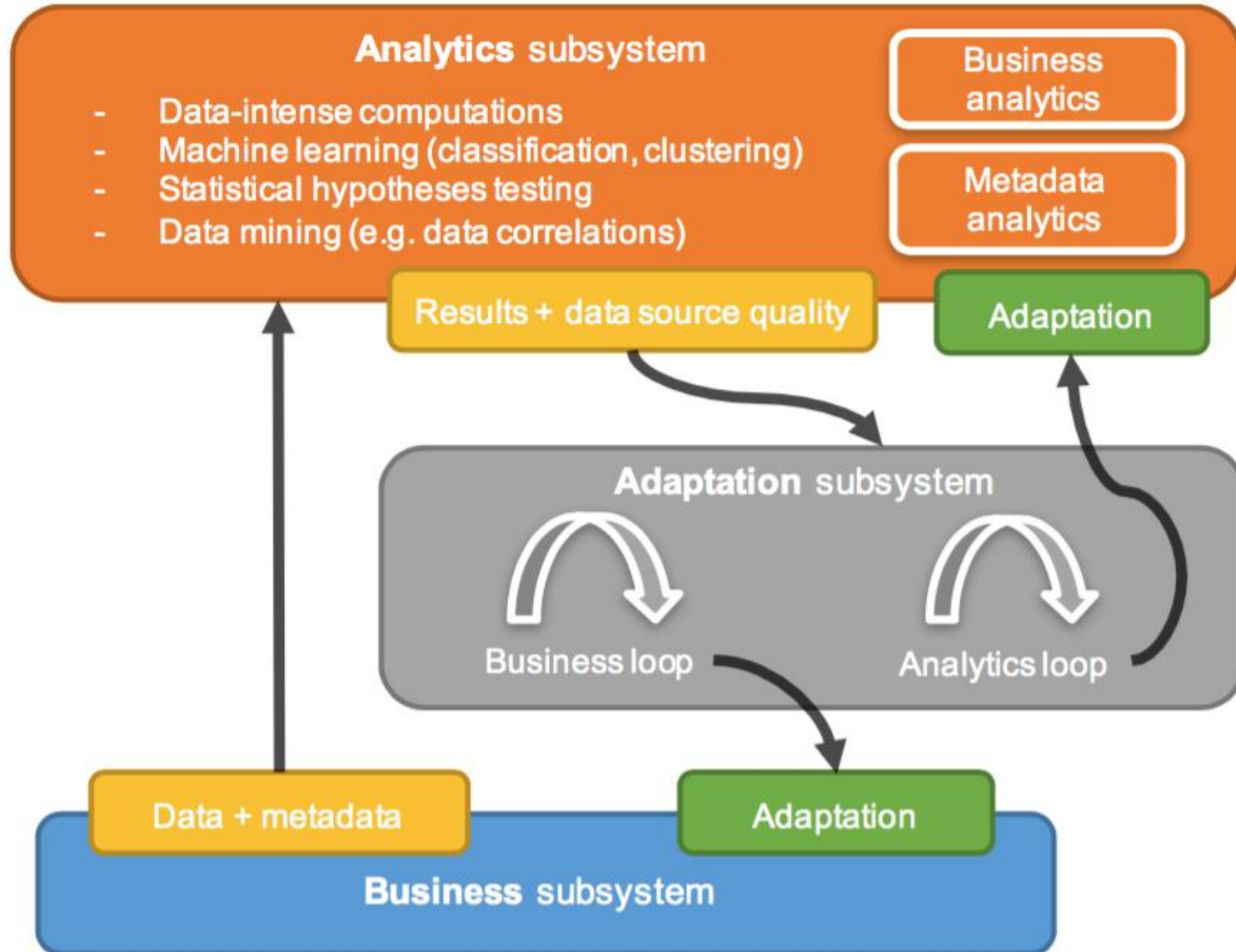
# Freeway Control System

- **All data gathered centrally to make control decisions**
- **Sensing devices are not always reliable**
  - Inaccuracy, faults, communication errors, environment changes...
- **System should acknowledge data veracity**
  - To estimate decision accuracy/confidence
  - To apply different weights to different data sources
  - To adjust/switch algorithms
  - All the above on runtime

# Outline

- Data-driven Systems
- Traffic Management Use Case
- **Theta Architecture**
- Theta in Action
- Discussion

# Θ Theta Architecture



# Θ Theta Architecture

## ■ **Business subsystem**

- Software-intensive cyberphysical system (+ external data)
- Controlled at runtime
- Sending out data + metada (accuracy, aggregation rates...)
- Receiving adaptation requests

## ■ **Analytics subsystem**

- Generate actionable information from data analysis
- Computations, statistical testing, machine learning, data mining
- Business analytics and Metadata analytics
- Receiving adaptation requests

## ■ **Adaptation subsystem**

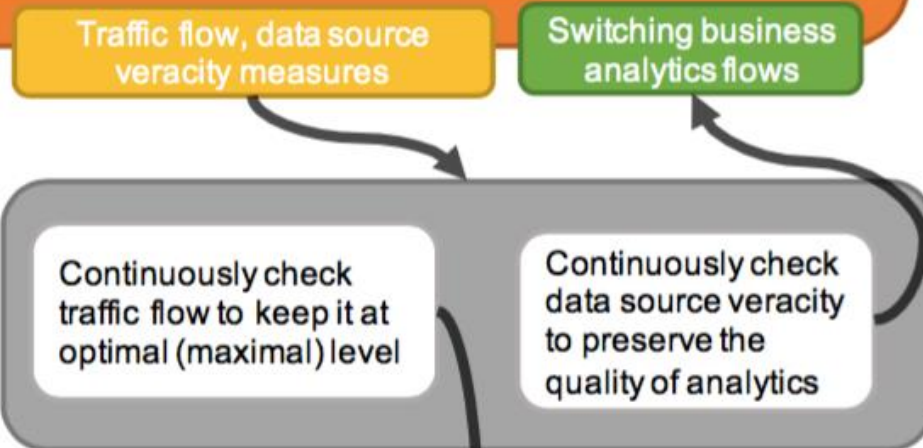
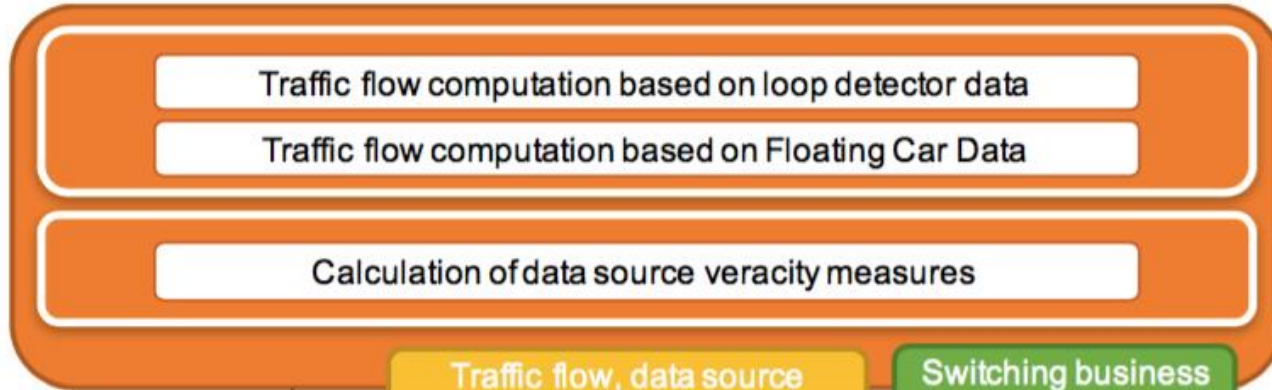
- Decision making based on received analytics and metadata
- Business and analytics loops as separate entities
- Reliable analytics to inform business decisions

# Outline

- Data-driven Systems
- Traffic Management Use Case
- Theta Architecture
- **Theta in Action**
- Discussion

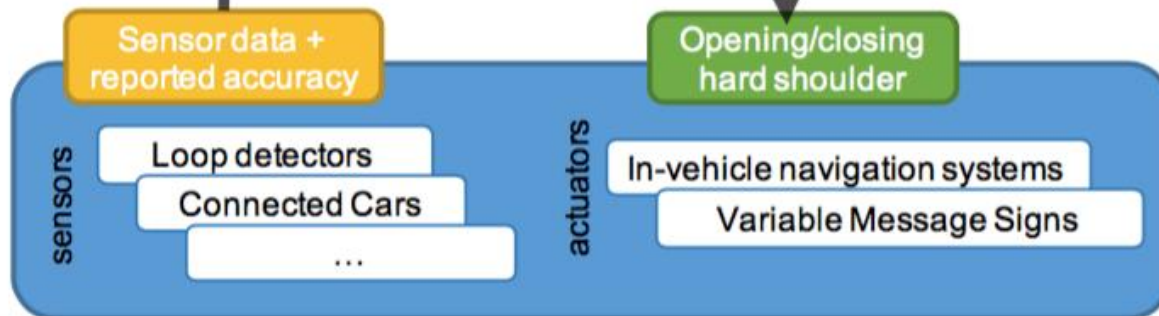
# Θ for Traffic Management

Analytics Subsystem



Adaptation Subsystem

Business Subsystem



# Outline

- Data-driven Systems
- Traffic Management Use Case
- Theta Architecture
- Theta in Action
- **Discussion**



# Observations

- **Theta = MAPE-K loop + Advanced Analytics**
- **No inference on loops complexity**
  - Each adaptation can be arbitrarily intricate
- **Ability to adapt between data sources**
  - Based on veracity, data privacy, confidentiality...

# Implementation Status

- **Current setting**
  - Testbed based on SUMO & python interface
  - TraCI: simulation of cars moving in a city
- **Analytics subsystem: Kafka**
- **Custom python engine for ML & data analysis**
- **Standard means from traffic engineering to assess veracity**
  - When veracity of speeds decreases, we switch to analytics based on loop detectors
- **Future work:**
  - Looking for ways to more accurately model and measure data veracity

Thank you!

