On-line time series classification:

Robust data representation for real-time data stream analysis

J. A. Barria

Intelligent Systems and Networks
Electrical and Electronic Engineering Department

Intelligent Systems & Networks

j.barria@imperial.ac.uk

Imperial College London

On-line time series classification

In collaboration with:

Dr. E.S. Garcia, Imperial College London, U.K.,
Dr. S. Thajchayapong, NECTEC, Thailand,
Dr. Q.R. Hamid, AkerSolutions, U.K.,
Dr. V. Alarcon-Aquino, UDLAP, Mexico.

1. Content

1. Internet of Things and Big Data

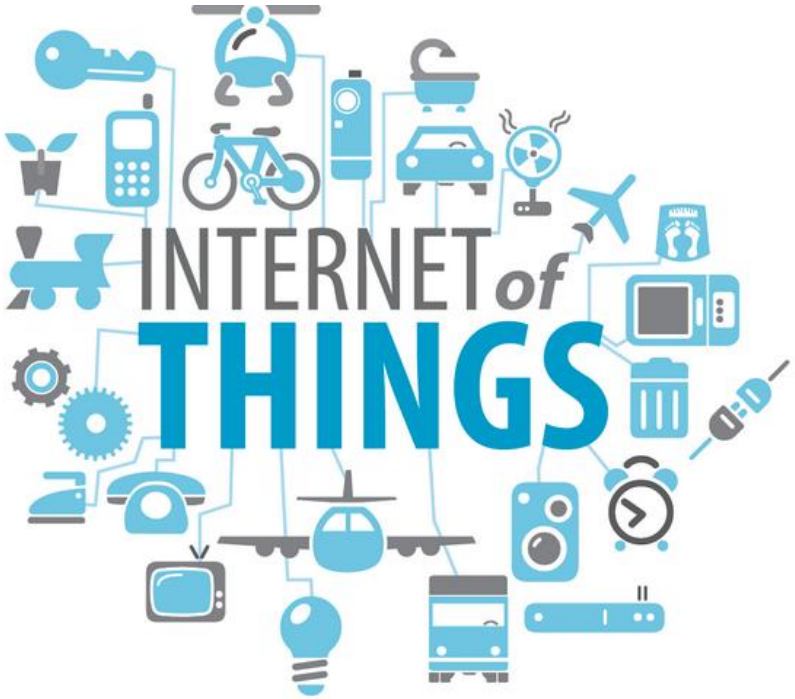
2. On-line Time Series Classification Framework
   - Pattern Recognition Approaches
   - Robust Data Representation
   - Novel Anomaly Detection Approach

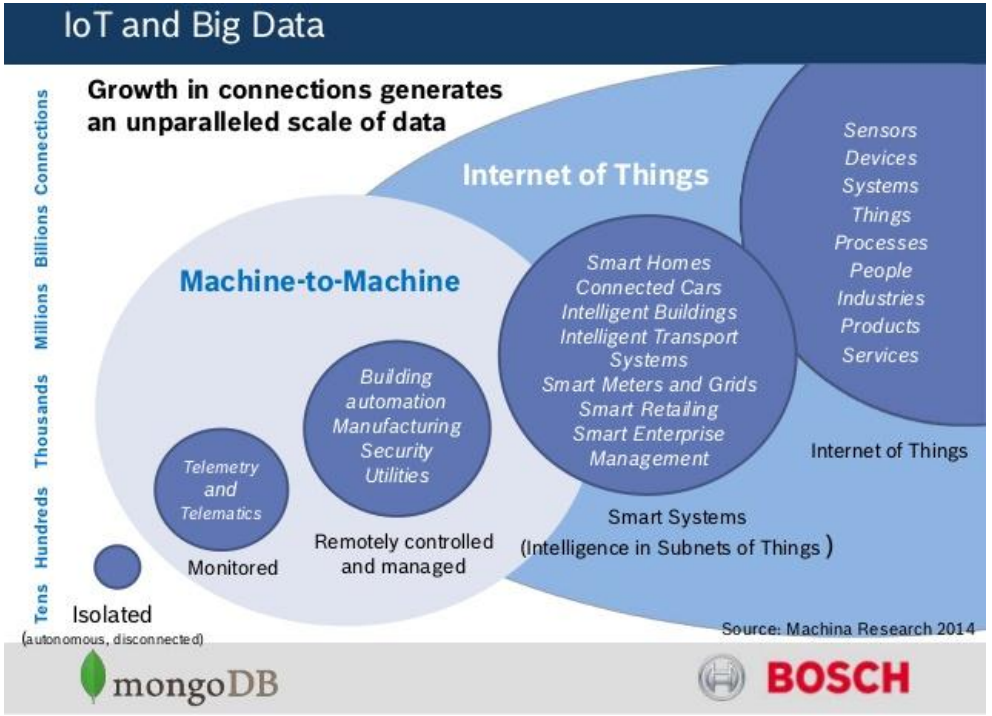
3. Real Time Monitoring Applications

4. Imperial College London

5. Final Remarks

1. The Internet of Things (IoT) and Big Data



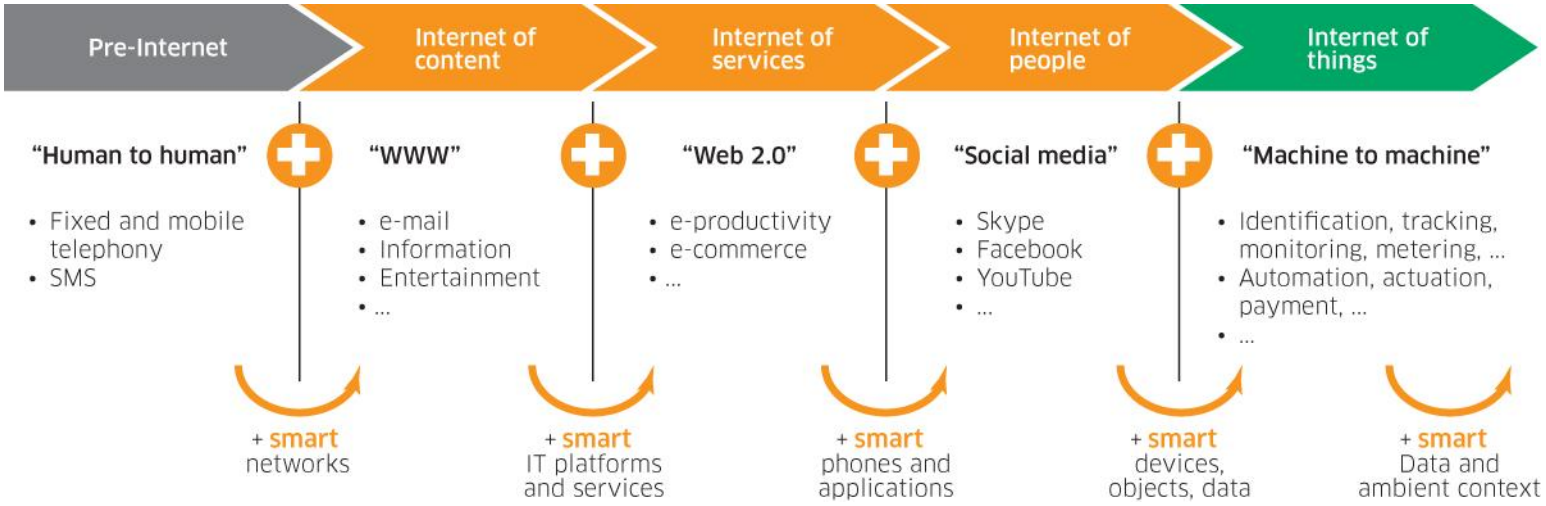


http://www.smartdatacollective.com

https://www.mongodb.com/presentations/

https://machinaresearch.com/

IoT and Big Data (2/10)



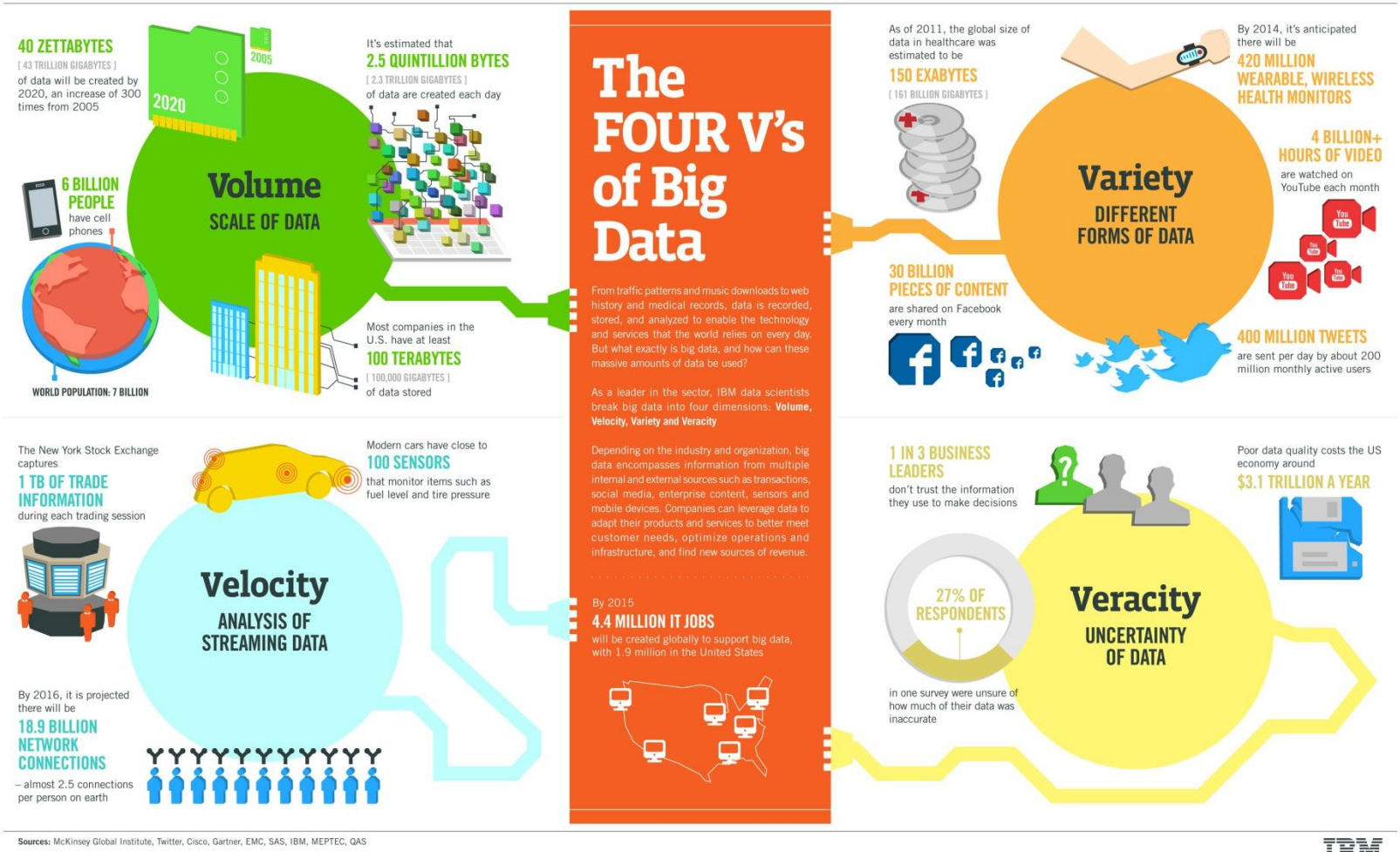
- Any application
- Any network
- Any service

https://techzine.alcatel-lucent.com/

Source: Alcatel-Lucent

j.barria@imperial.ac.uk

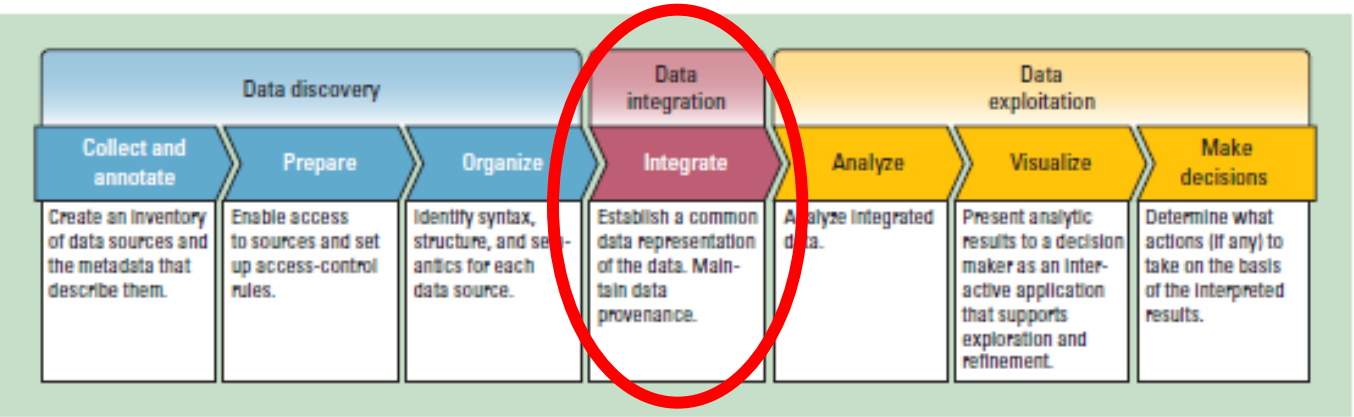
Big Data: the Four V's (3/10)



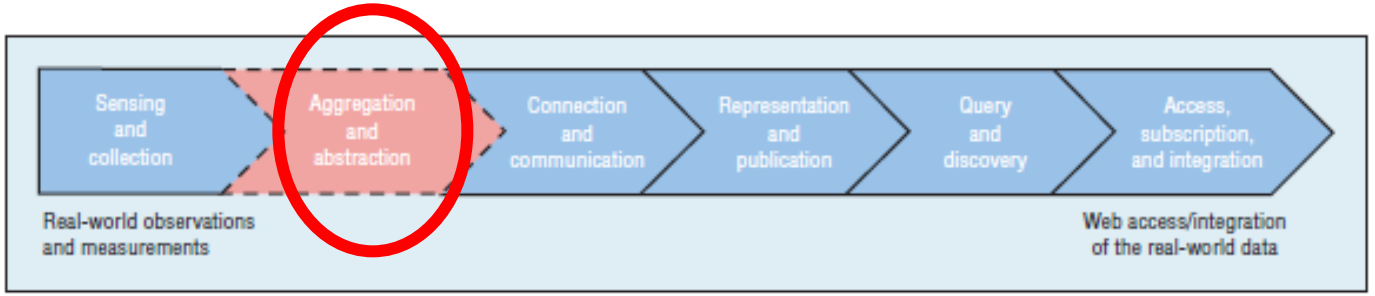
Source: http://www.ibmbigdatahub.com/infographic/four-vs-big-data

j.barria@imperial.ac.uk

Data Representation (4/10)

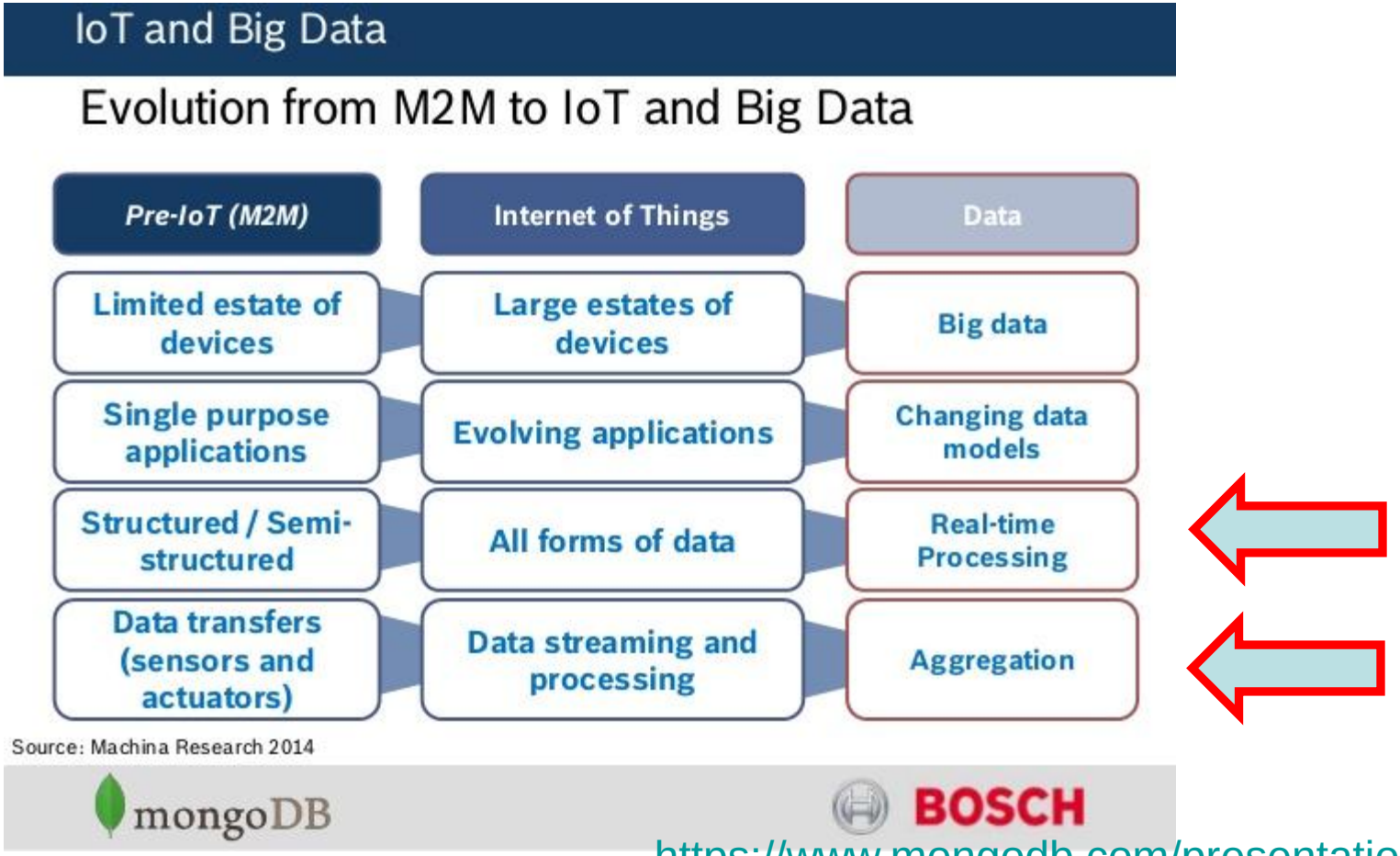


Re.: H. G. Miller and P. Mork, “From Data to Decisions: A Value Chain for Big Data,” It Pro., v. 15, no. 1, pp. 57–59, 2013.



Re.: P. Barnaghi, A. Sheth, and C. Henson, "From Data to Actionable Knowledge: Big Data Challenges in the Web of Things [Guest Editors' Introduction]," Intelligent Systems, IEEE, v.28, no.6, pp.6,11, Nov.-Dec. 2013

From M2M to IoT and Big Data (5/10)



https://www.mongodb.com/presentations/

https://machinaresearch.com/

j.barria@imperial.ac.uk

Big Data Management (6/10)

- From important competency to a critical differentiation: From Volume & Variety to Velocity.
- Big Data Technology: Operational vs. Analytical.
  - Operational features: Real-time & interactive intelligence – low latency, on-line capture.
  - Analytical features: Complex analysis – high throughput. http://www.mongodb.com
- Combine Operational and Analytical technologies.

➢ Generic/global fast solutions rather that application specific solutions.

j.barria@imperial.ac.uk

Data Streaming: Volume, Variety and Velocity(*) (7/10)

- Support a wide variety of data(*)
- Focus on agility and flexibility rather than conformity(*)

➢ Domain Independent framework: builds the foundation to unsupervised classification algorithms.

Source: **(*) http:** //www-935.ibm.com/services/us/gbs/thoughtleadership/2014analytics/

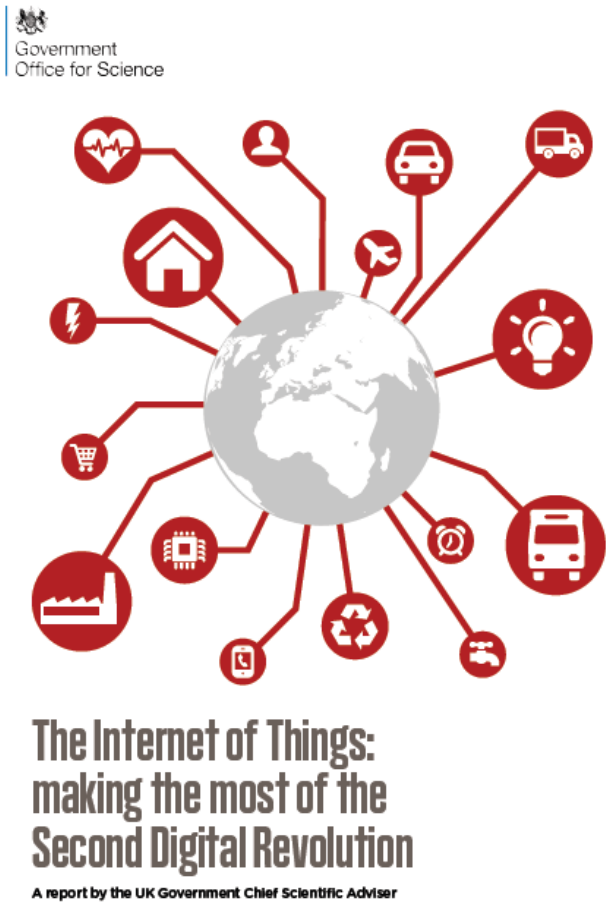
Data Streaming: Volume, Variety and Velocity(*) (8/10)

- Foster rapid data consumption(*)
- Provide prompt access to relevant information(*)
- Make pervasive use of predictive analytics a priority (*).

➢ Real-time/on-line time series classification problems.

Source: **(*) http:** //www-935.ibm.com/services/us/gbs/thoughtleadership/2014analytics/

Applying the Internet of Things (9/11)

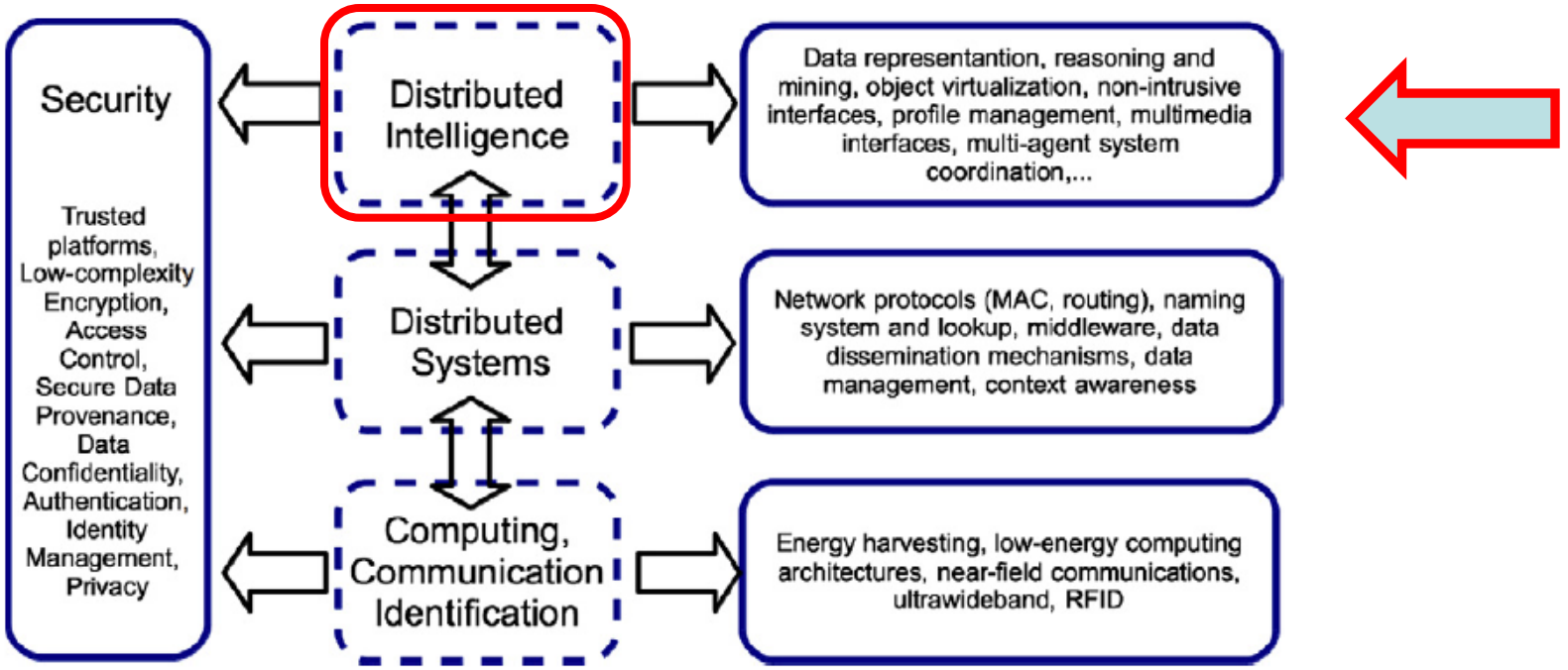


https://www.gov.uk

- Transport
- Energy
- Healthcare
- Agriculture
- Buildings

j.barria@imperial.ac.uk

IoT Relevant Research areas (10/10)



(D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, “Internet of things: Vision, applications and research challenges,” Ad Hoc Netw., vol. 10, no. 7, pp. 1497–1516, 2012.)

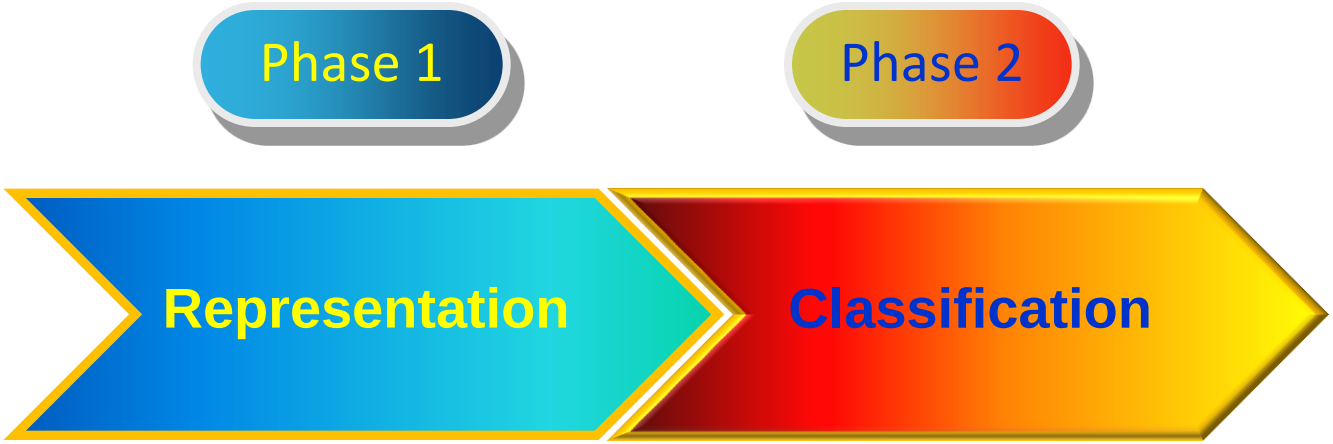
2. On-line Time Series Classification Framework

One important on-line Big Data application is: Real-time monitoring & anomaly detection.

- *Health condition monitoring* as well as *anomaly detection* are classification problems.
- They are time series classification problems because they involve the *time element* (e.g. short-term and long-term condition changes).
- *Pattern recognition* frameworks belongs to the field of classification methods.

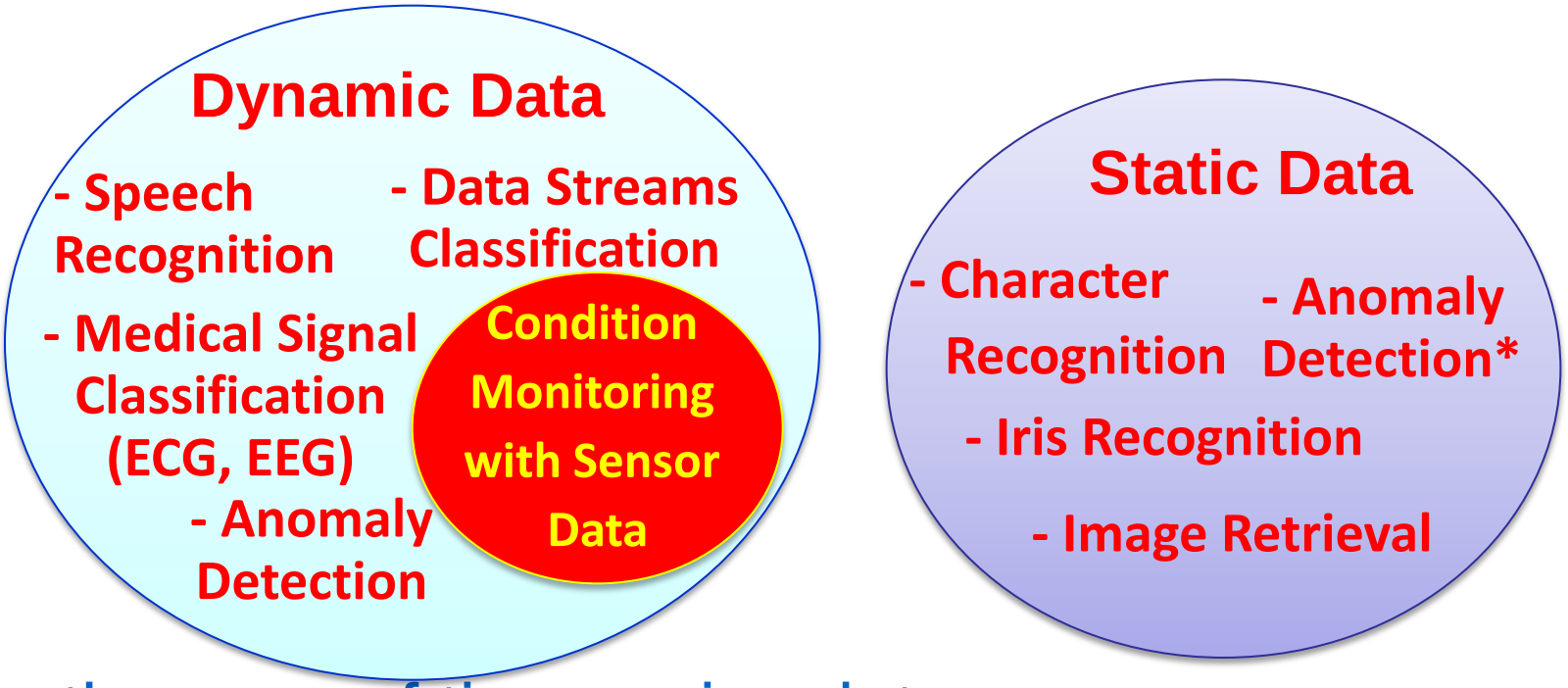
2.1. The Pattern Recognition (PR) Approaches (1/5)

- Composed of two phases:



- Extracting appropriate properties or features.
- Mapping the features to a particular class.

Different Data type: Different Applications (2/5)

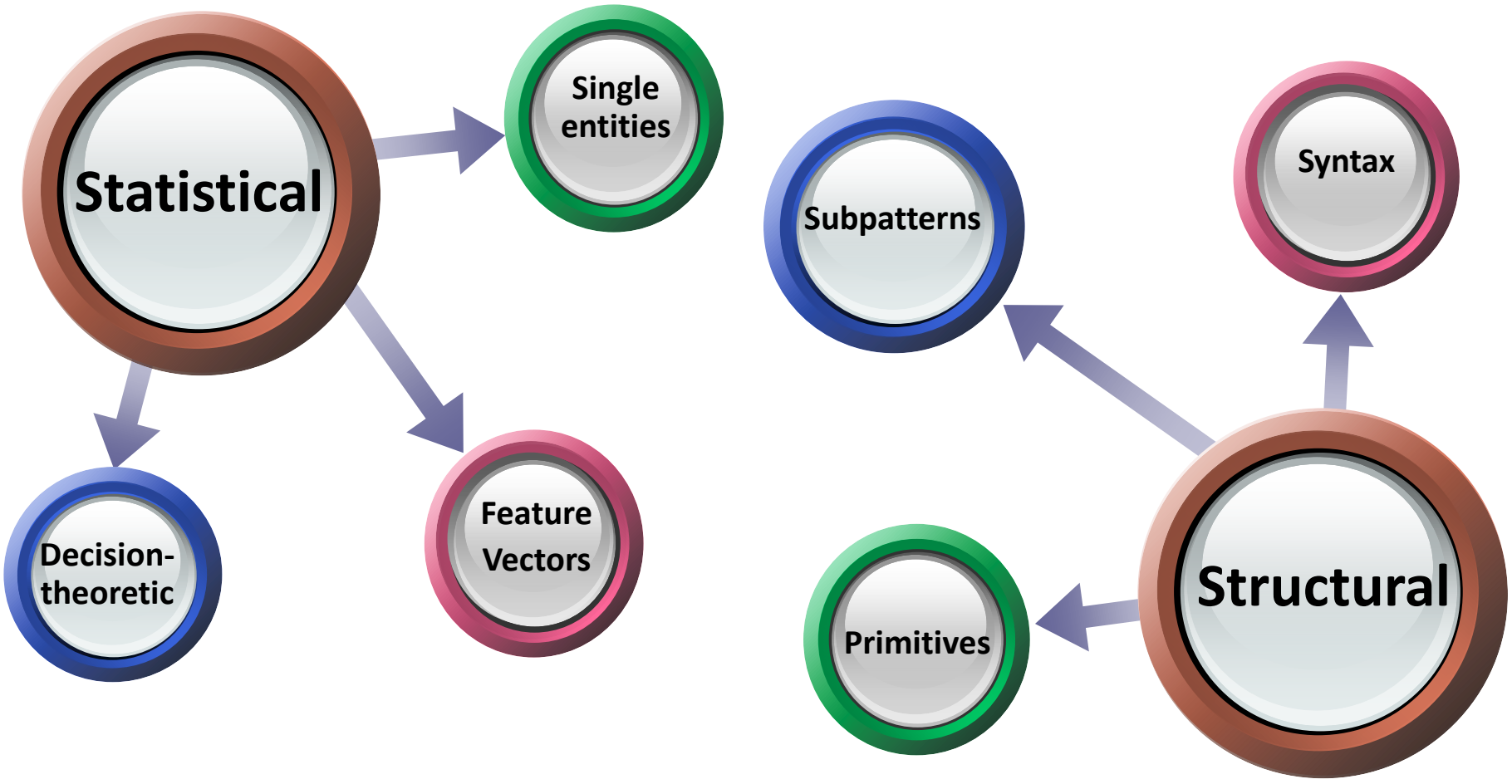


For the case of time series data:

- A major concern is finding *appropriate pattern representations*.
- Most of the approaches treat time series as static data (hence they do not offer suitable representations).

j.barria@imperial.ac.uk

Pattern Recognition Paradigms (3/5)



Pattern Recognition Paradigms (4/5)

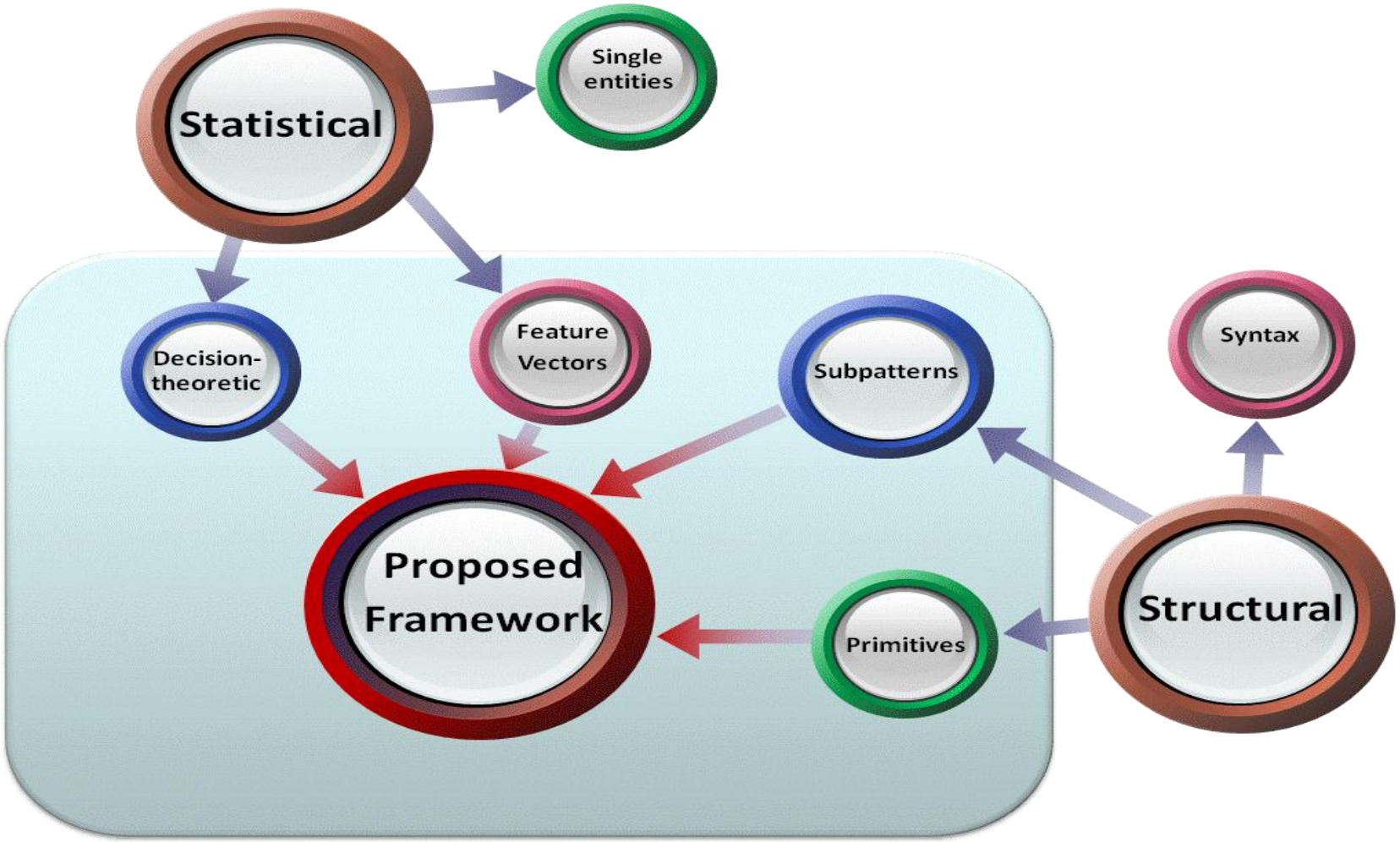
Statistical:

- Patterns treated as single entities and described by numerical feature vectors.
- The classification involves the partition of the feature space into regions.

Structural:

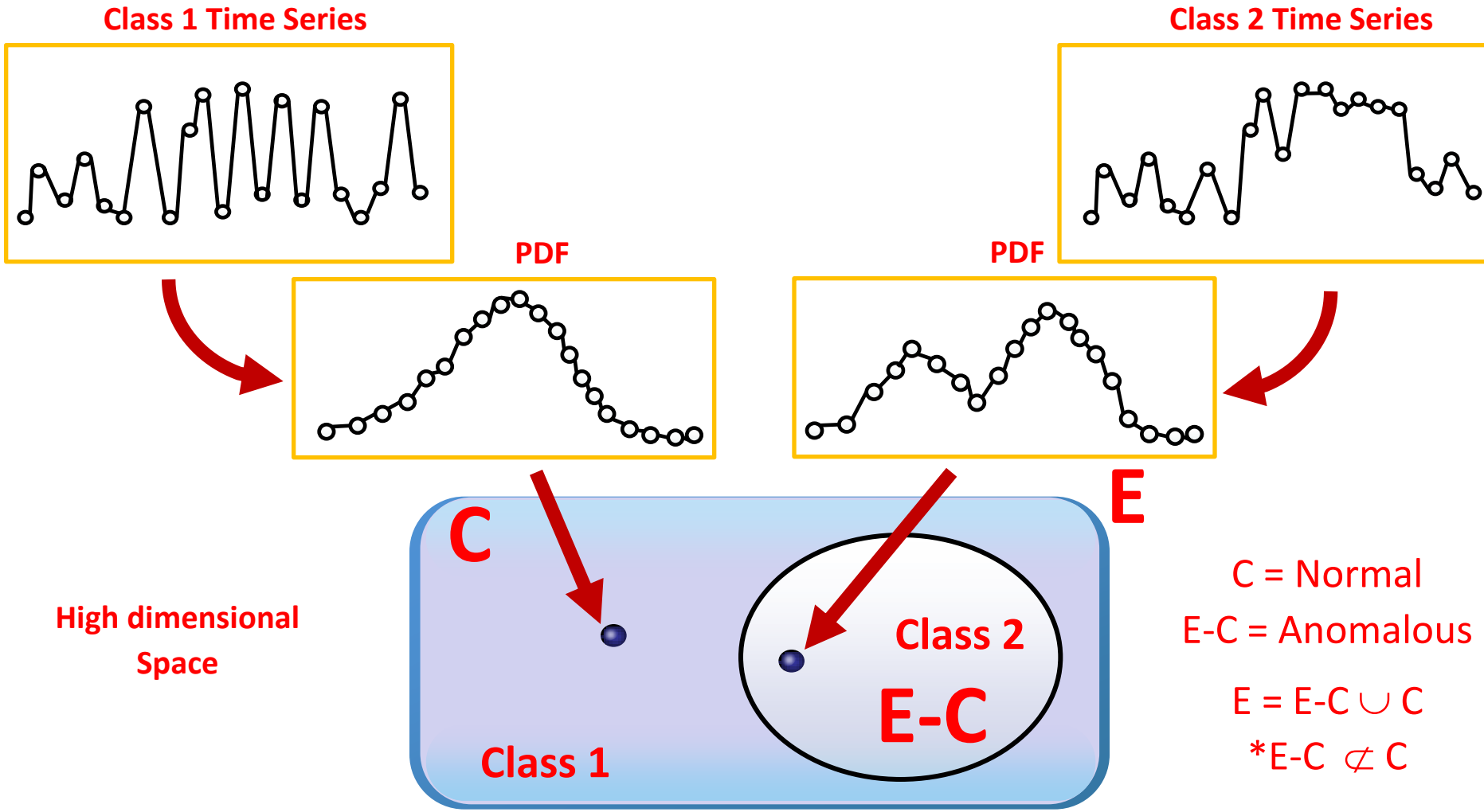
- Patterns treated as combination of multiple entities and described by its topological relations.
- The classification involves matching the structural representation according to a syntax.

Proposed Approach: New PR Representation Framework (5/5)



j.barria@imperial.ac.uk

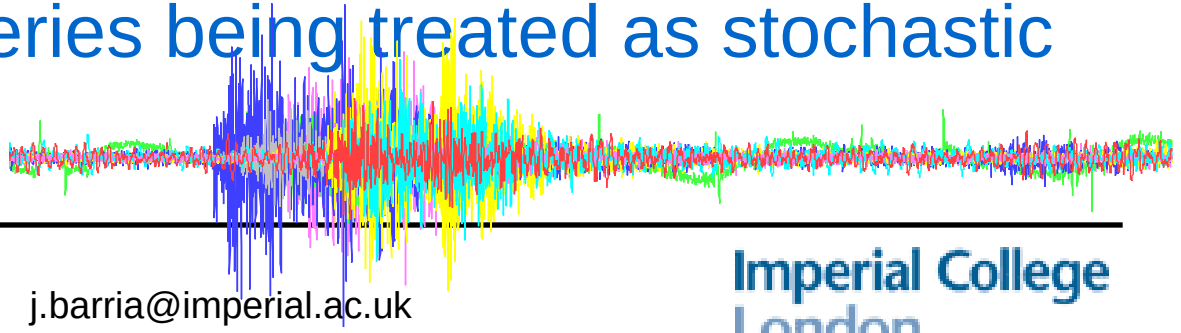
2.2. Robust Data Representation (1/2)



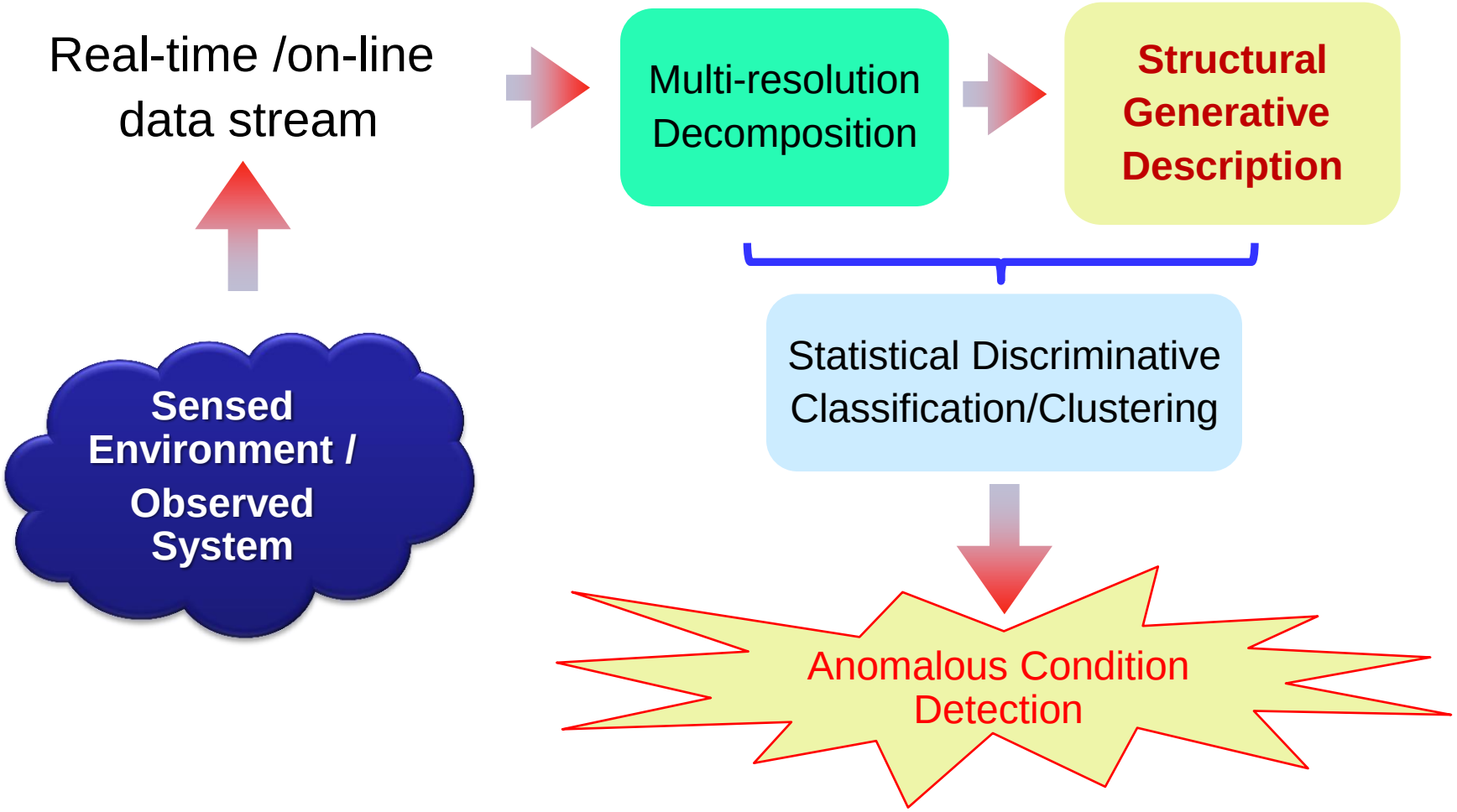
j.barria@imperial.ac.uk
Imperial College London

Structural Generative Description (SGD): The Proposed Approach (2/2)

- **Domain independent** : which makes the framework suitable for a wide variety of applications.
- It is an on-line / real-time approach.
- Constructs robust time series represenations.
- Based on time series being treated as stochastic processes.

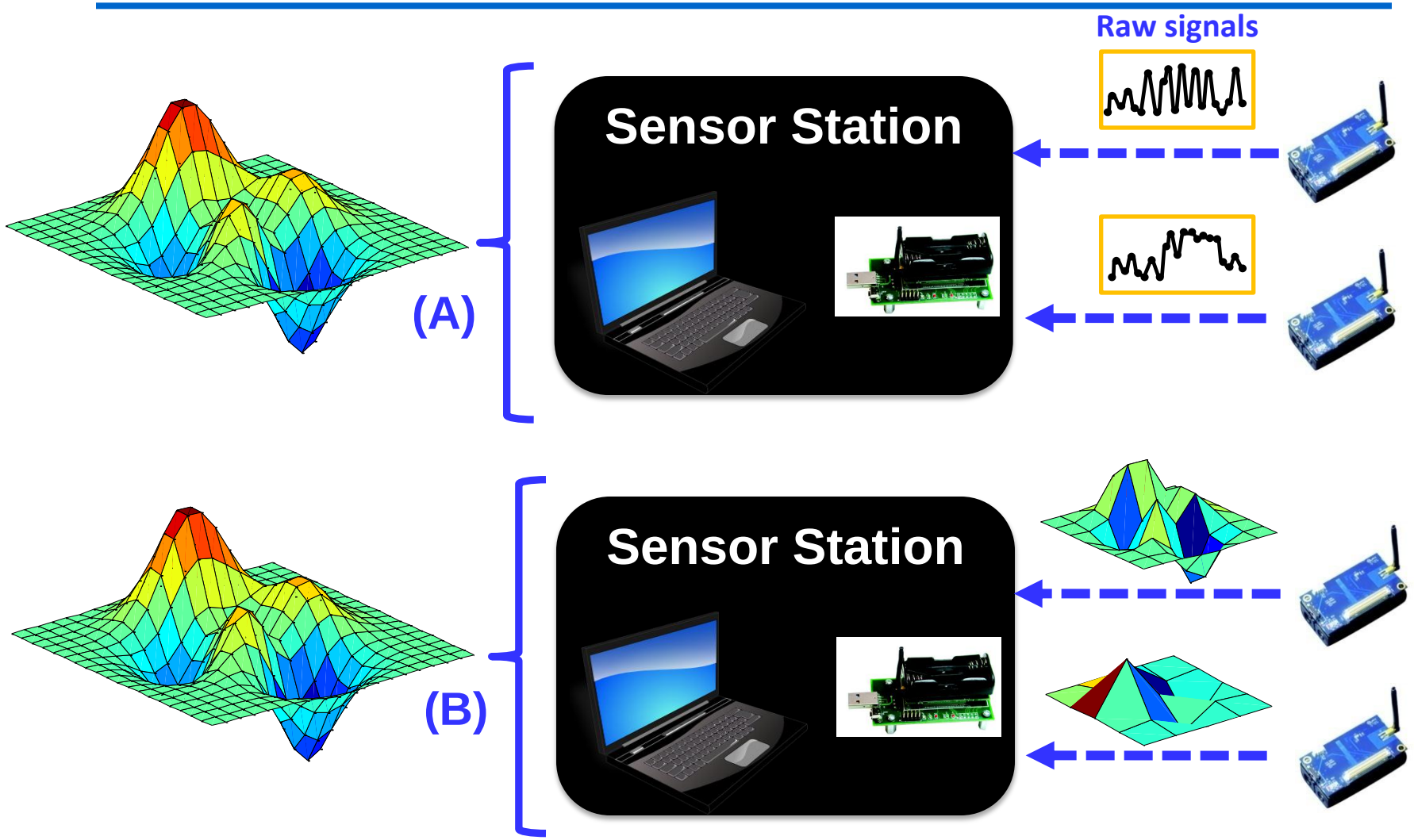


2.3. Novel Anomaly Detection Approach (1/2)



j.barria@imperial.ac.uk

Implementation: Centralised (A) vs Distributed (B) (2/2)



j.barria@imperial.ac.uk

3. Real-time Monitoring: Applications

On-line time series classification framework and its applications:

3.1. Biometric Recognition and Forensics,

3.2.i. Smart Infrastructures Monitoring,
3.2.ii. Machine/motor Health Conditioning Monitoring,

3.3. Distributed Charging of EVs,

3.4.i. Transportation Networks Monitoring,
3.4.ii. Environmental (pollution) Monitoring.

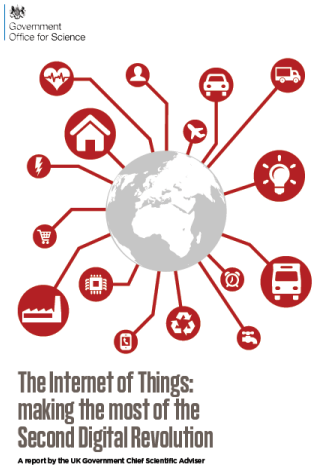
j.barria@imperial.ac.uk

3.1. Applying the Internet of Things: Healthcare

- Healthcare

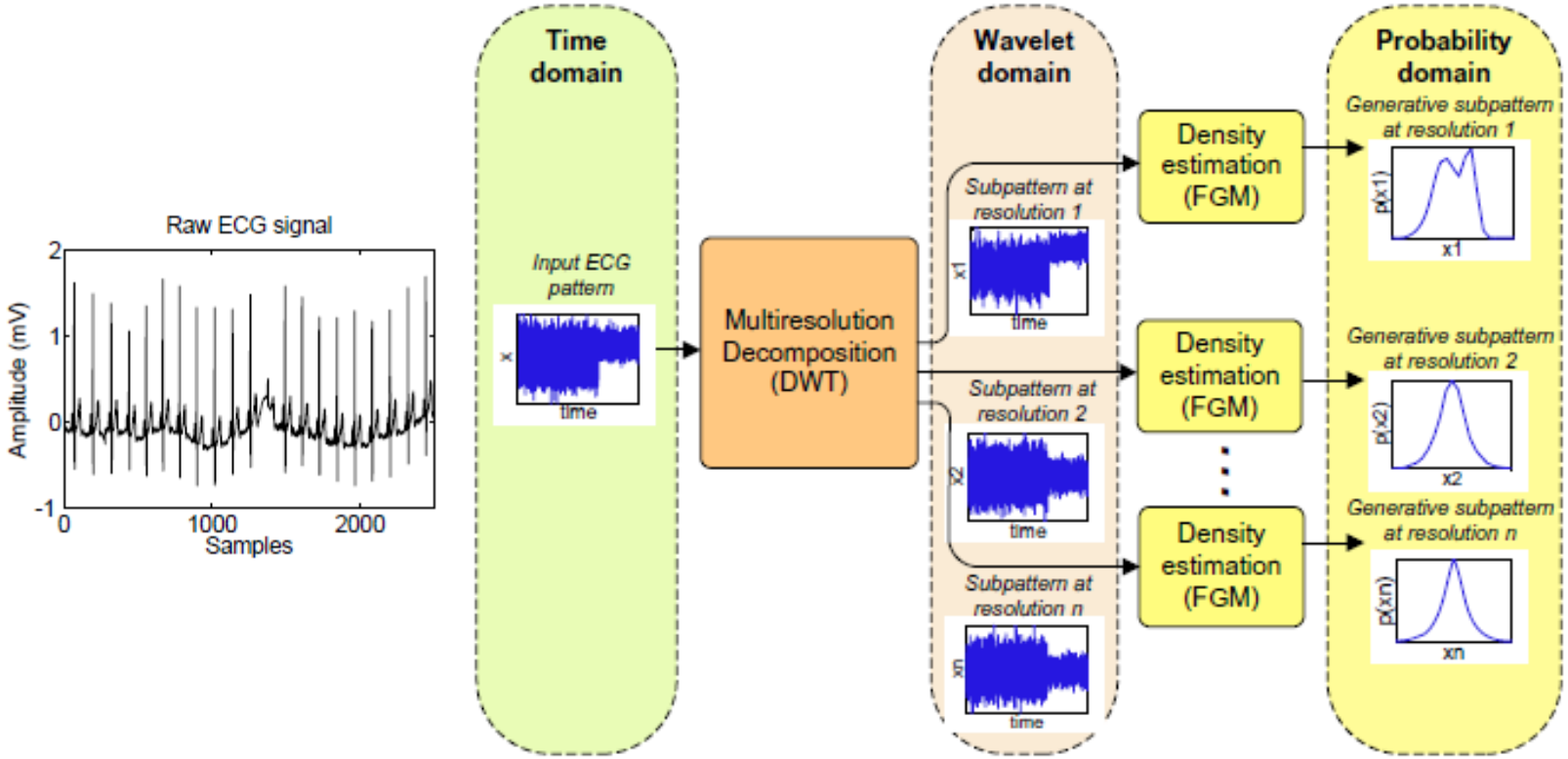
➢Prevention and early identification,

- Research,
- Data security and ownership,
- Hardware security and interoperability,
- Change management.



https://www.gov.uk

Biometric Recognition and Forensics (1/2)



j.barria@imperial.ac.uk

Biometric Recognition and Forensics (2/2)

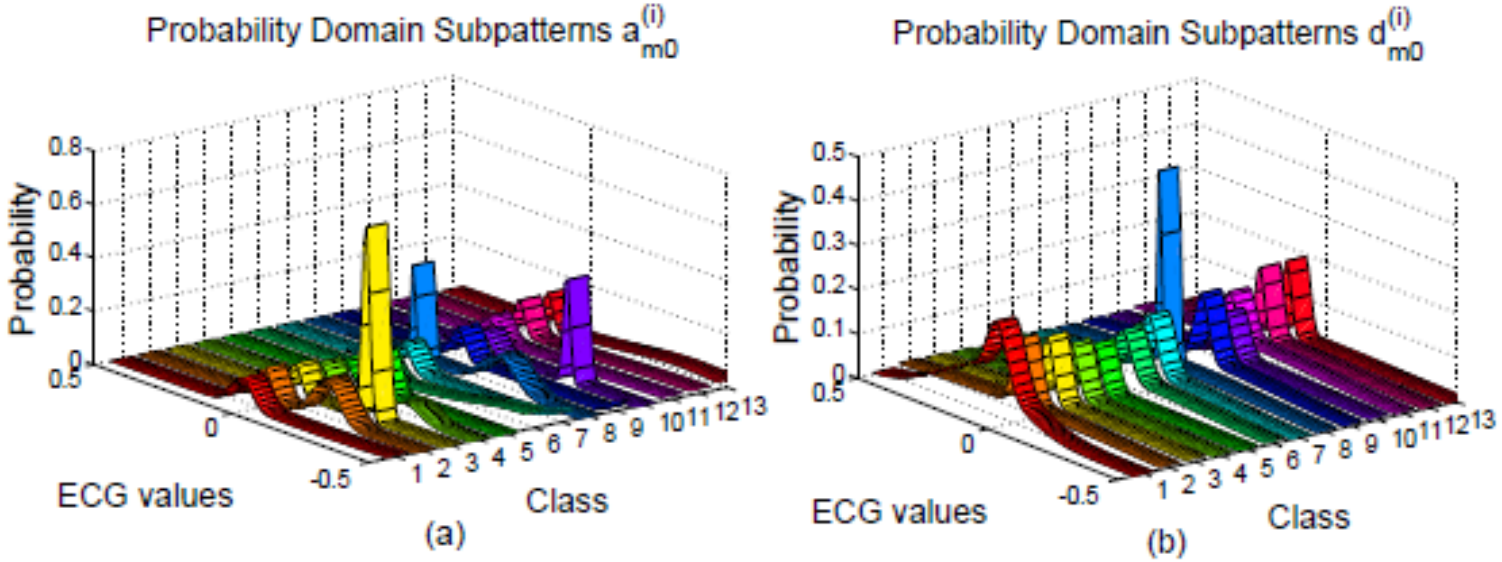


Fig. 5: Probability domain subpatterns for ECG data windows from the thirteen classes with *db*3, $W = 12$, $M = 1$, $G = 6$, $I = 2^9$, $T = 2^{-4}$, $C = 2^6$, $\lambda = 2^{-3}$. (a) Scaling coefficients; (b) Detail coefficients.

j.barria@imperial.ac.uk

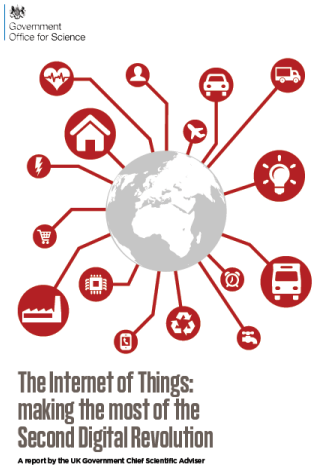
3.2. Applying the Internet of Things: Buildings

- Buildings

  ➢Optimising design and minimising costs,

  - Increasing comfort,

  ➢Security and safety.

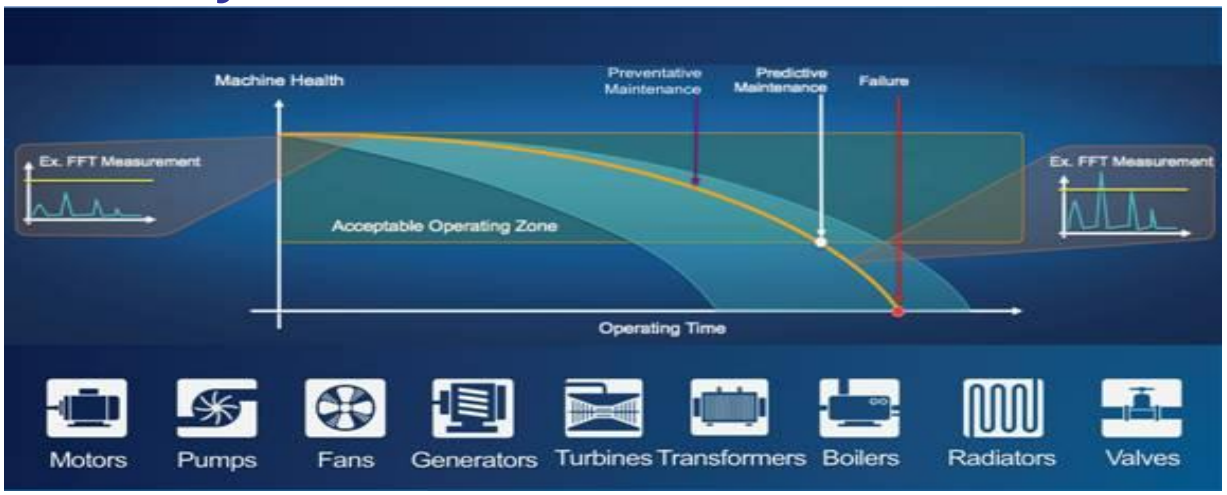


https://www.gov.uk

j.barria@imperial.ac.uk

I. Smart Infrastructures: SI (1/3)

- Real time assessment of the deterioration of civil infrastructure (i.e. bridges, tunnels, water pipelines).
- Framework based on WSN, condition monitoring, and anomaly detection.



(http://www.ni.com/newsletter/52418/en/)

- Proactive classification into e.g. adequate condition, maintenance required, and in need of replacement.

j.barria@imperial.ac.uk

SI: Machine Health Conditioning Monitoring (2/3)

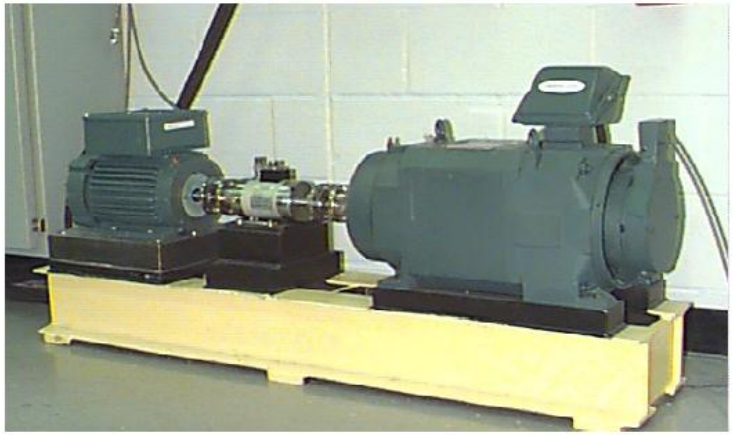


Fig. 3: Test stand for Experiment 1 [68]

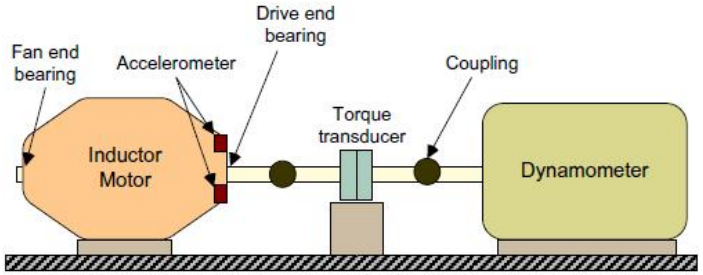
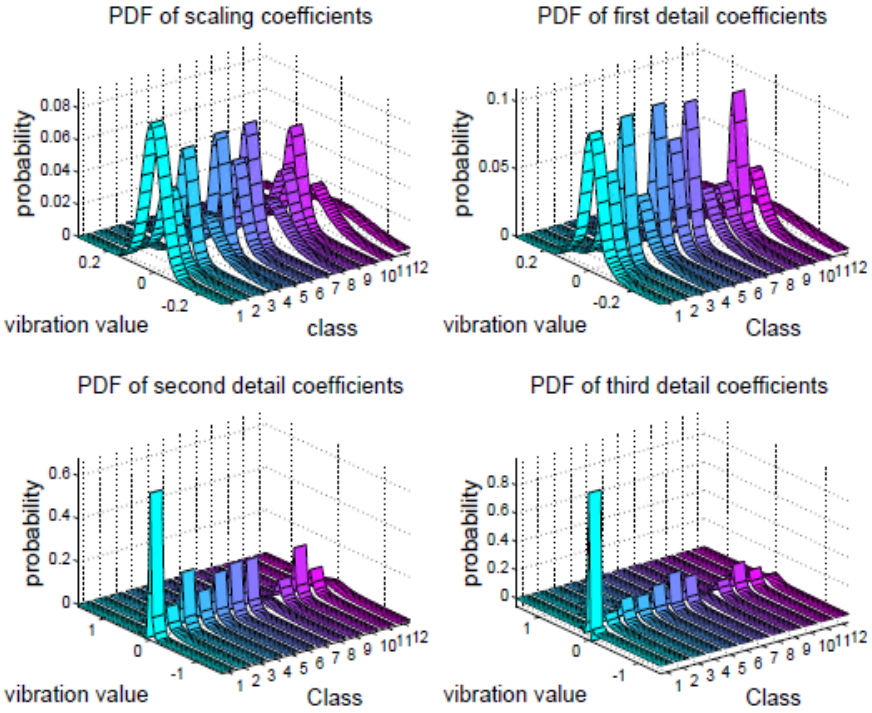
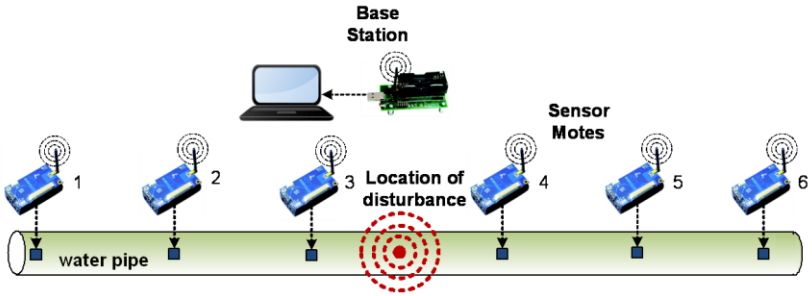


Fig. 4: Scheme diagram of test stand for Experiment 1 [66]

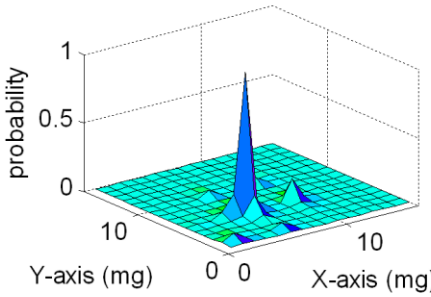


(http://www.eecs.case.edu/laboratory/bearing/)

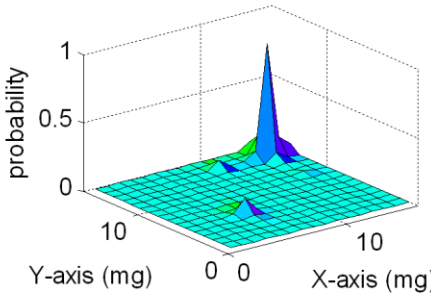
SI: Sensor Network Resilience: Design (3/3)



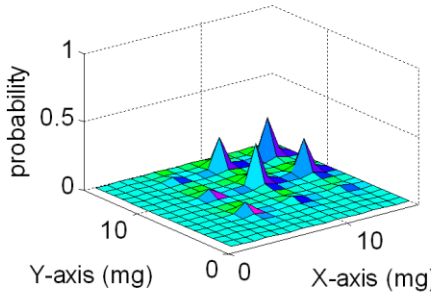
PDF Sensor ID:01 Vibration: Normal



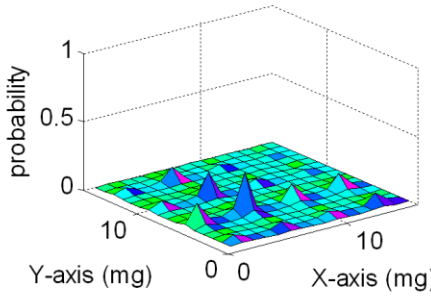
PDF Sensor ID:01 Vibration: Pattern 1

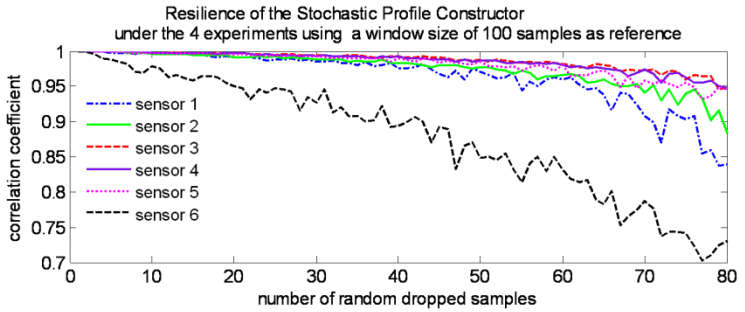


PDF Sensor ID:01 Vibration: Pattern 2



PDF Sensor ID:01 Vibration: Pattern 3

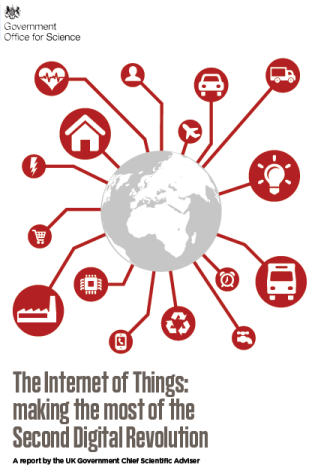




j.barria@imperial.ac.uk

3.3. Applying the Internet of Things: Energy

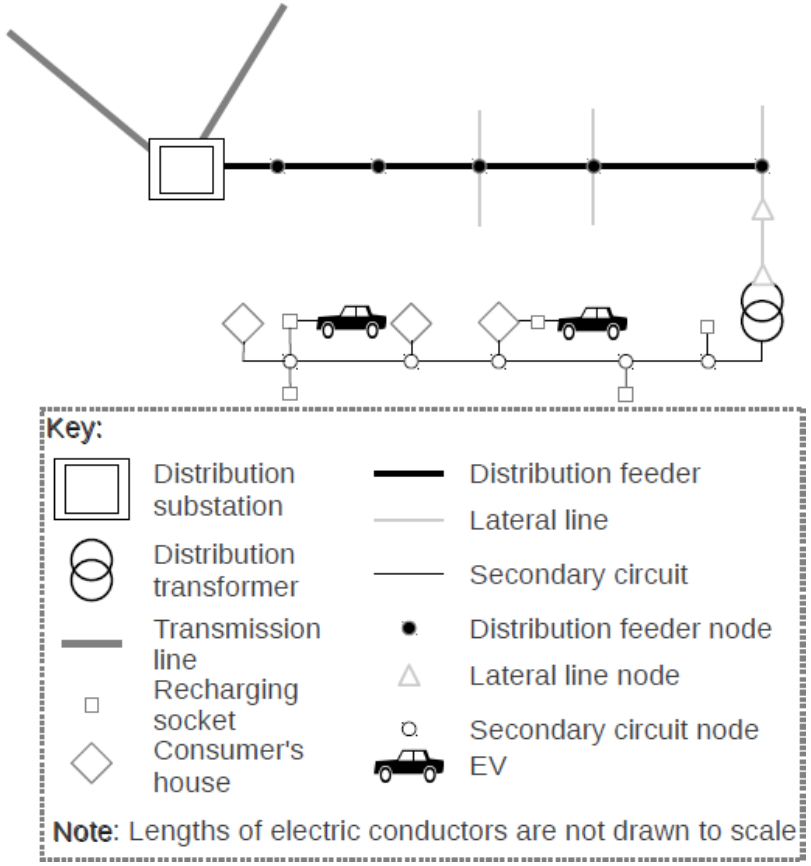
- Energy
  - Reducing energy demand,
  - ➢Managing energy patterns,
  - Driving innovation,
  - Increased energy demand,
  - Security and standards,
  - Variable access.



https://www.gov.uk

EVs participation for Frequency Regulation Service (1/3)

Power Distribution Network Scenario



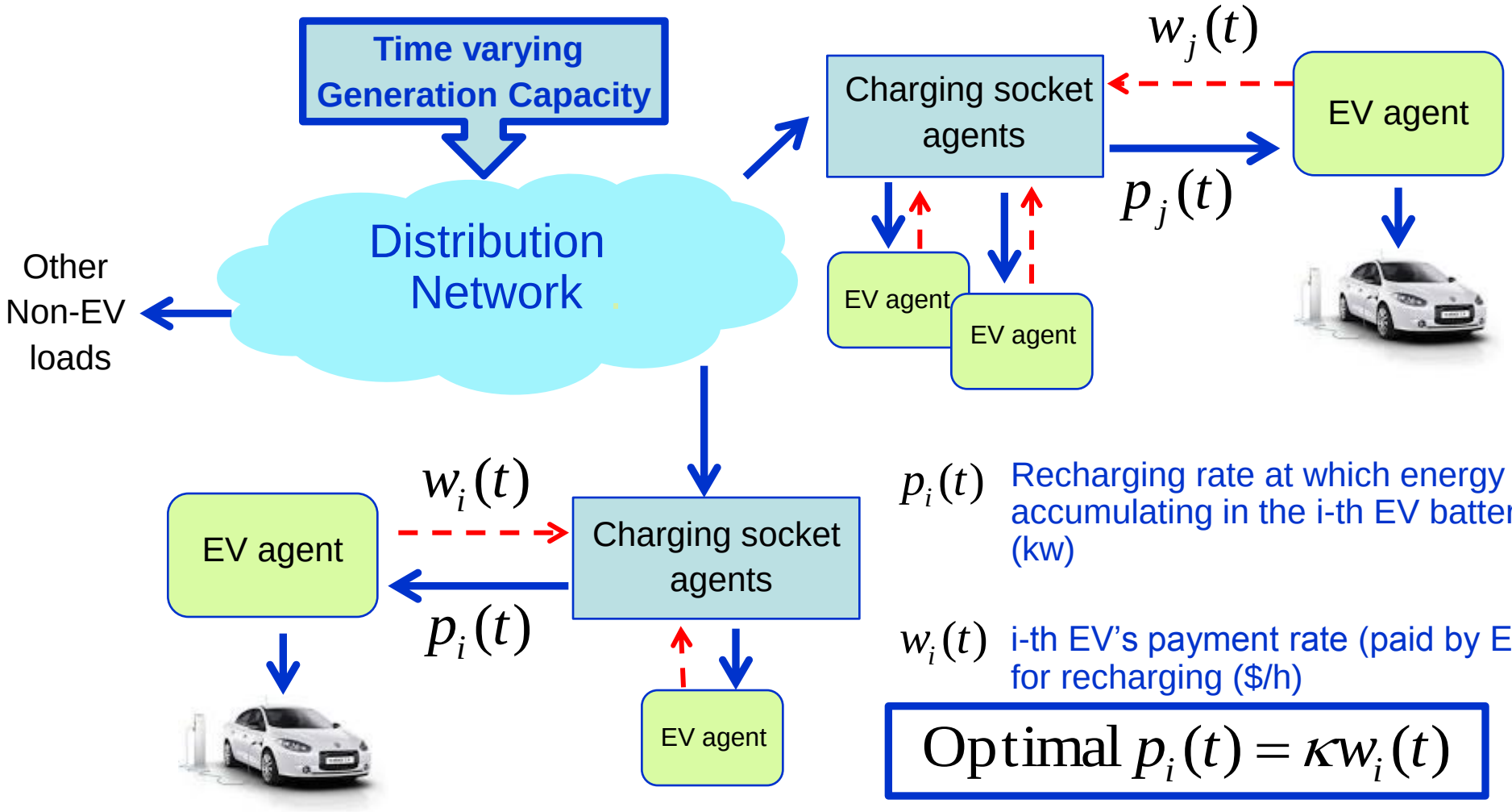
- Two mechanism to facilitate EVs participation in operational aspects of a smart grid.
  - Frequency regulation,
  - Congestion avoidance.
- The solution needs to be distributed and dynamic.

Intelligent Systems & Networks

j.barria@imperial.ac.uk

Imperial College London

Frequency Regulation Service (2/3)



$p_i(t)$ Recharging rate at which energy is accumulating in the i-th EV battery (kw)

$w_i(t)$ i-th EV's payment rate (paid by EV) for recharging ($/h)

$$\text{Optimal } p_i(t) = \kappa w_i(t)$$

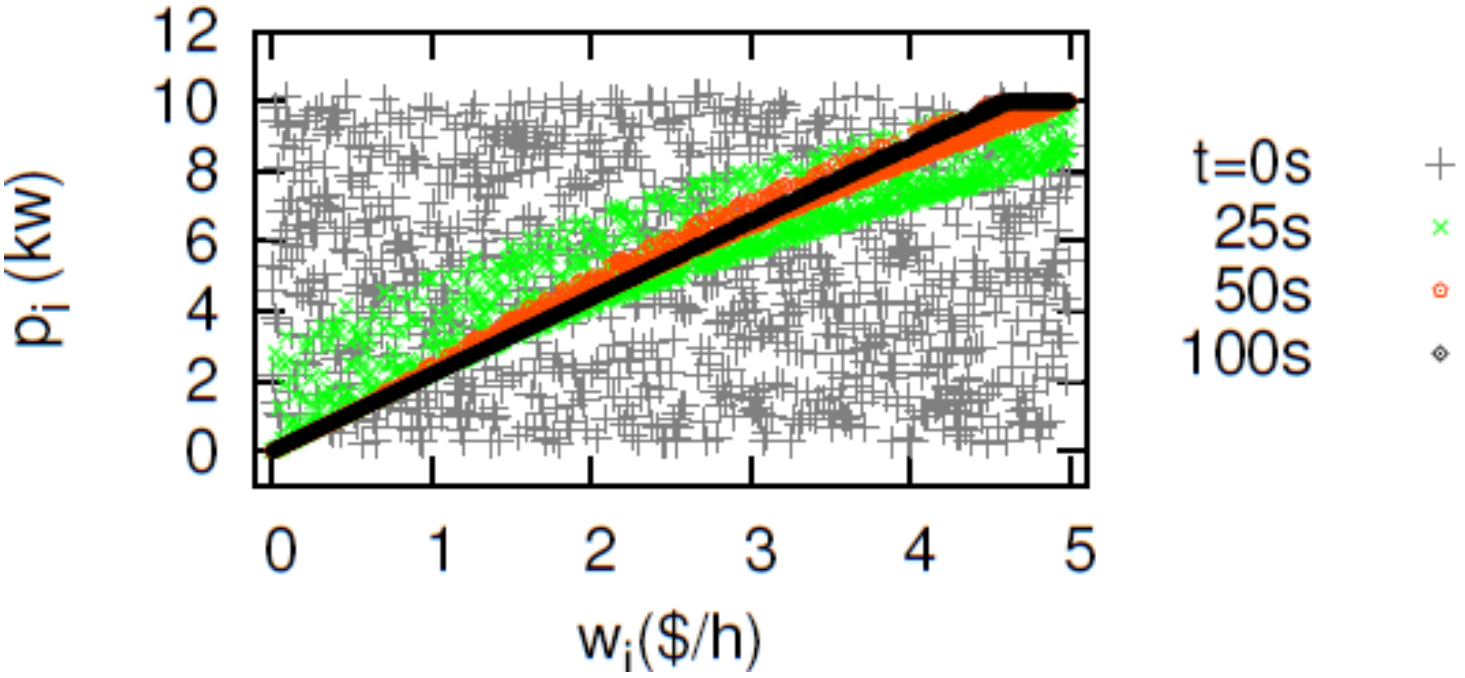
Intelligent Systems & Networks

j.barria@imperial.ac.uk

Imperial College London

Convergence of Recharging rates of 1000 EVs (3/3)

$$p_i(t) = \kappa w_i(t)$$

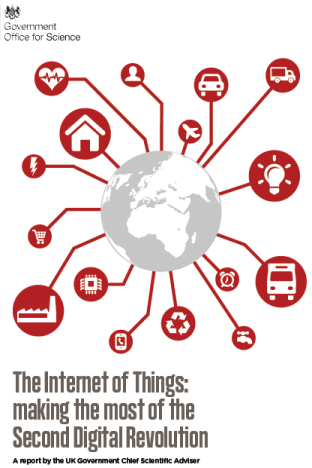


Note: The initial recharging rates $p_i(t)$ are randomly assigned.

j.barria@imperial.ac.uk

3.4. Applying the Internet of Things: Transport

- Transport
  - Passenger journeys,
  - ➢Increased safety,
  - Transporting goods,
  - Security, reliability and regulation.



https://www.gov.uk

j.barria@imperial.ac.uk

I. Transportation Networks: ITS (1/2)

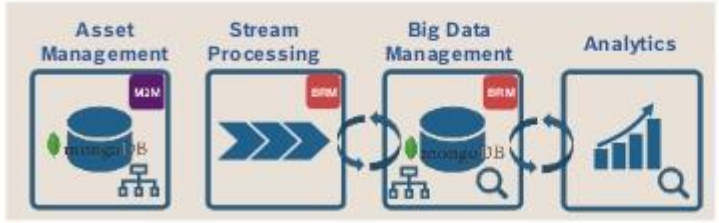
- Management Tools for Safe Mobility
- Classification of Vehicular Traffic Anomalies
- Spatial Inference of Traffic Conditions

j.barria@imperial.ac.uk

Transport vehicles: ITS (2/2)

IoT and Big Data

Use Case 3: Field Data Capturing



Project SCFD

- Structured Capturing of Field Data
- Components: Car brakes, power steering, etc.
- Usage patterns: temperature, voltage, etc.
- Predictive maintenance, product optimization

Why MongoDB:

- Constantly evolving system, from a data capturing and a data analytics point of view
- Large amount of streaming data

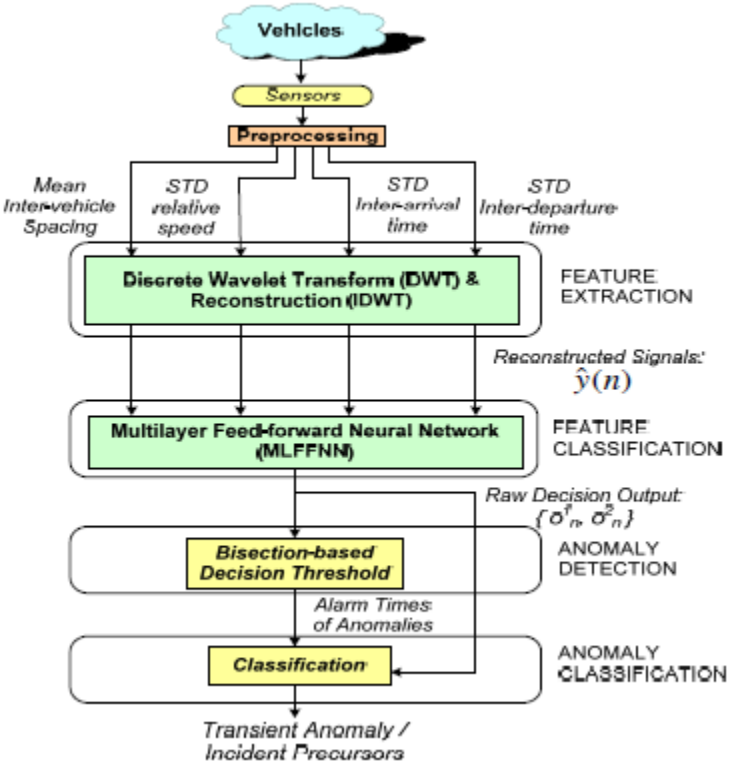
mongoDB BOSCH

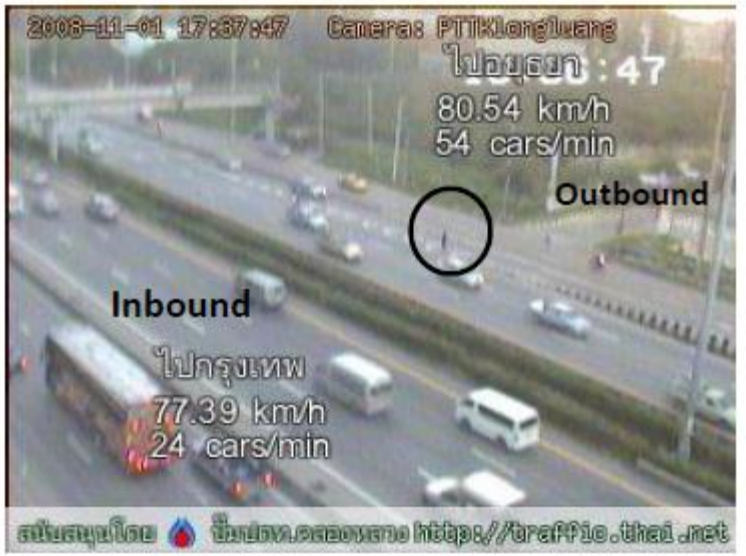
https://www.mongodb.com/presentations/

https://machinaresearch.com/

j.barria@imperial.ac.uk

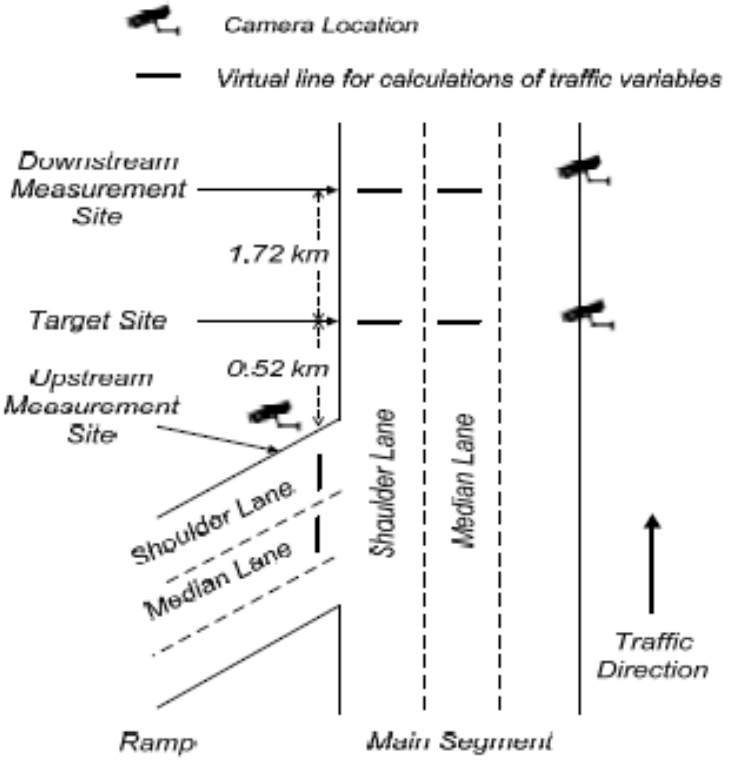
ITS: Distributed Classification of Vehicular Traffic Anomalies (1/2)

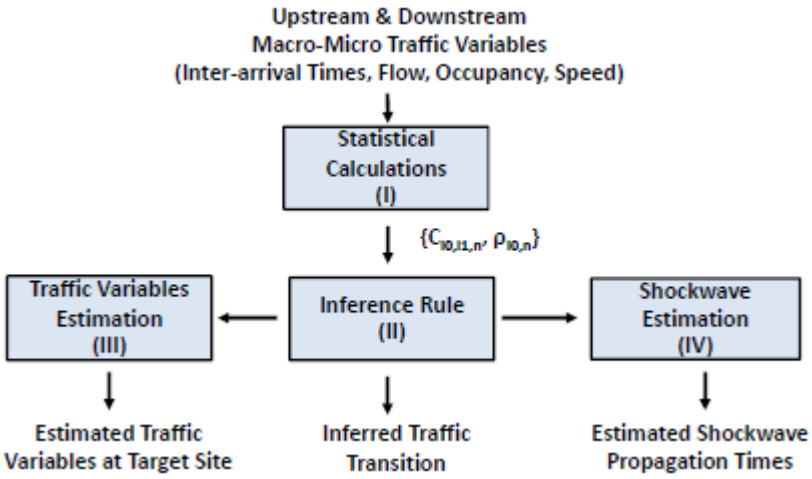




j.barria@imperial.ac.uk

ITS: Spatial Inference of Traffic Transitions (2/2)





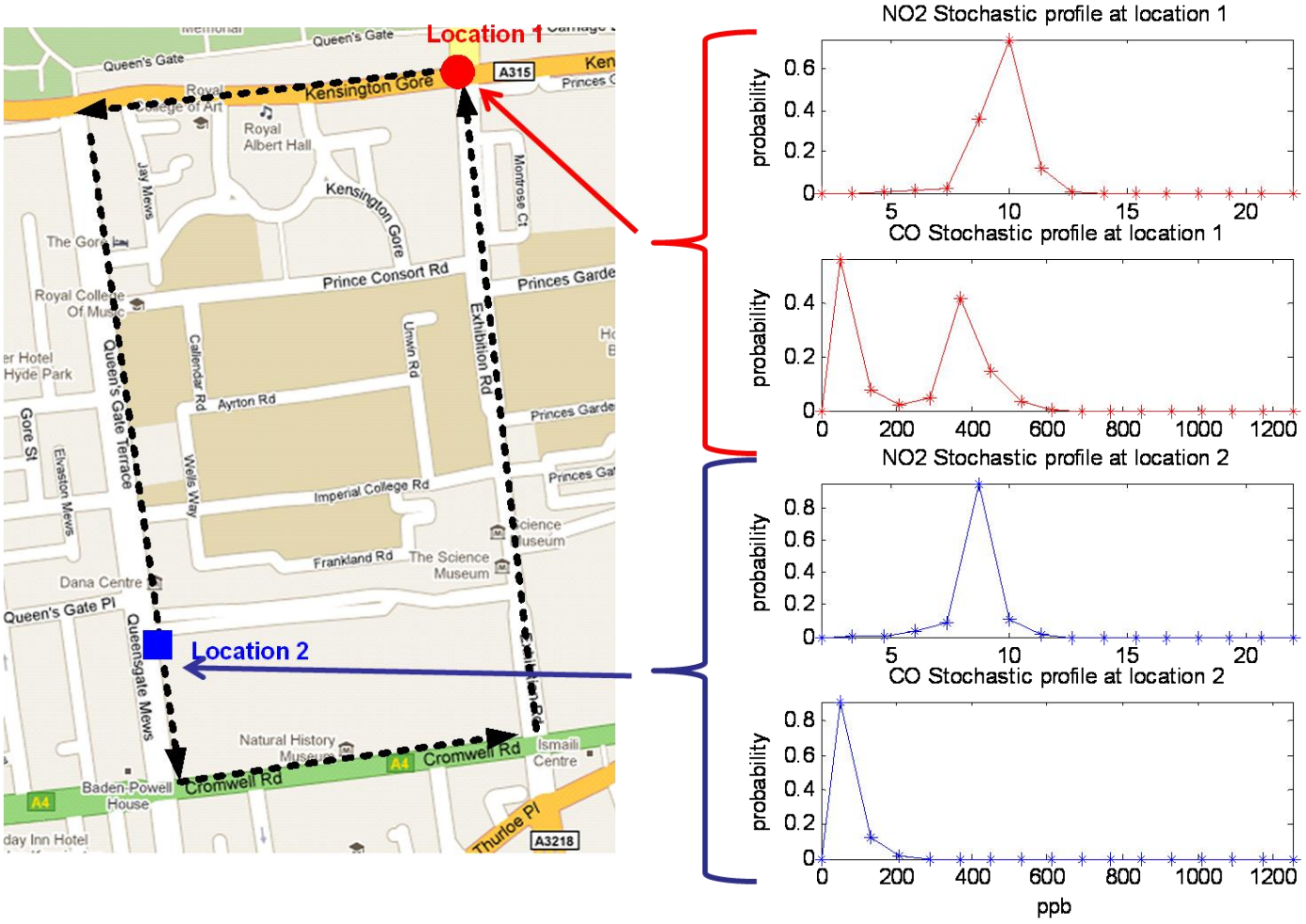
j.barria@imperial.ac.uk

II. Real time Environmental Monitoring (1/3)

- Monitoring of vehicle(s) behaviour and real-time construction of environmental pollution patterns map.
- Intelligent monitoring systems for real-time characterisation and construction of pollution level maps.

j.barria@imperial.ac.uk

Real time Construction of Pollutant levels (pdf) (2/3)

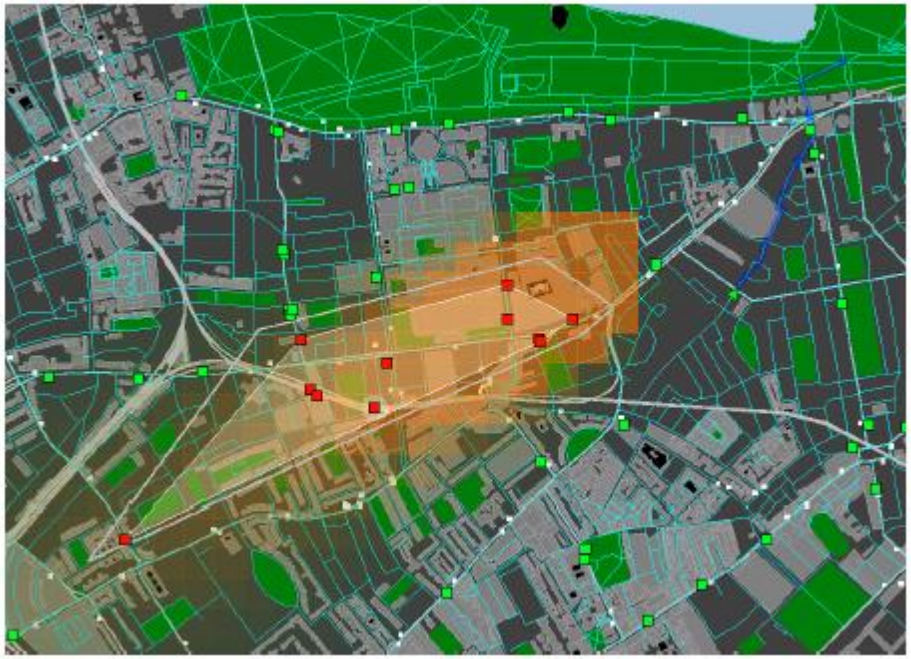


Intelligent Systems & Networks

j.barria@imperial.ac.uk

Imperial College London

Participatory Sensing Mobility Models (3/3)



- Urban Environment.
- Mission-oriented, participatory sensor network deployment.
- Combining fixed and mobile sensor nodes.

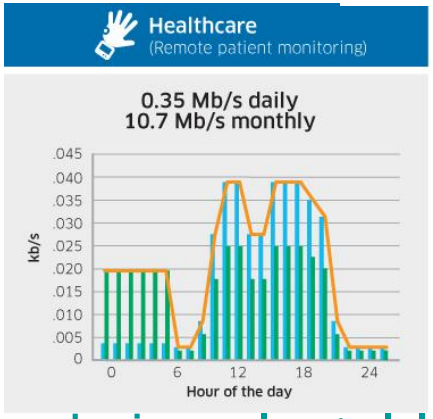
j.barria@imperial.ac.uk

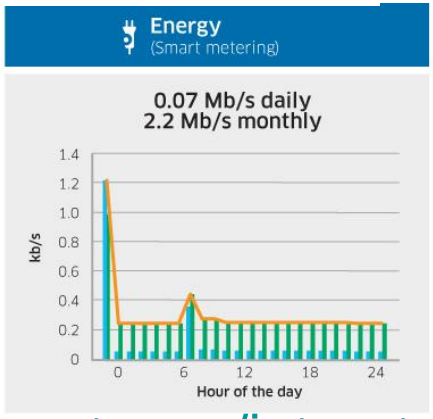
Interoperability across networks: new traffic patterns

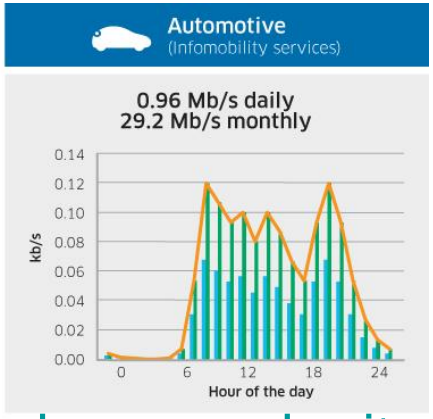
- Any service
- Any application
- Any network

- Applications
- Connectivity
- Devices
- Sensors

Source: Alcatel-Lucent





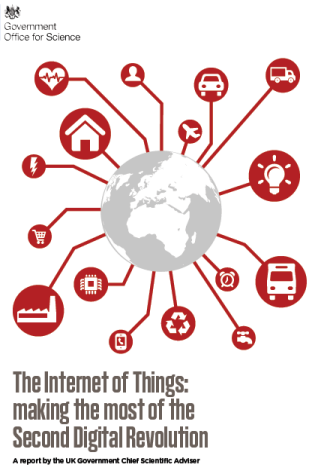


https://techzine.alcatel-lucent.com/iot-network-can-make-it-or-break-it

j.barria@imperial.ac.uk

3.5. Applying the Internet of Things: Agriculture

- Agriculture
  - Maximising yield,
  - Improving food traceability,
  - Tackling environmental challenges,
  - Incompatibility,
  - Lack of infrastructure,
  - Technical expertise.



https://www.gov.uk

4. Imperial College London, July 2014

https://www.imperial.ac.uk

Imperial College London

5. Final Remarks

From competency to a critical differentiation:
From Volume & Variety to Velocity.

- Big Data Technology: Operational vs. Analytical.
- Combine Operational and Analytical technologies.
- Generic/global fast solutions rather that application specific solutions.

- Domain Independent framework.
- Real-time time series classification problems.
- Foundation to unsupervised classification algorithms.

j.barria@imperial.ac.uk

References

1. S. Thajchayapong and J. A. Barria, “Spatial Inference of Traffic Transition using Micro-Macro Traffic Variables”, IEEE Transactions on Transactions on Intelligent Transportation Systems, 16 (2) pp. 854 - 864 (2015).
2. E. S. Garcia-Trevino and J.A. Barria, “Structural Generative Descriptions for Time Series Classification”, IEEE Transaction on Cybernetics, 44 (10) pp. 1978 - 1991 (2014).
3. S. Thajchayapong, E. S. Garcia-Trevino and J. A. Barria, “Distributed Classification of Traffic Anomalies using Microscopic Traffic Variables”, IEEE Transactions on Transactions on Intelligent Transportation Systems, 14 (1) pp. 448 - 458 (2013)
4. E. S. Garcia-Trevino and J. A. Barria, “A Statistical-Structural Framework for ECG Biometric Recognition” submitted 2015
5. E. Garcia-Trevino, J. A. Barria, “Online wavelet-based density estimation for non stationary streaming data”, Computational Statistics and Data Analysis, 56 pp. 327–344 (2011).
6. V. Alarcon-Aquino and J.A. Barria, Change Detection in Time Series using The Maximal Overlap Discrete Wavelet Transform, International Journal LAAR 39 (2), pp. 145-152 (2009).
7. V. Alarcon-Aquino and J.A. Barria, “Multi-resolution FIR Neural Network Based Learning Algorithm Applied to Network Traffic Prediction” IEEE Transactions on Systems, Man, and Cybernetics – C, 36 (2), (2006) pp. 208-220.

References

8. V. Alarcon-Aquino and J.A. Barria, “Anomaly Detection in Communication Networks using Wavelets”, IEE Proc.- Comm., 148 6, (2001), pp. 355-362.
9. J. A. Barria and S. Thajchayapong, “Detection and Classification of Traffic Anomalies using Microscopic Traffic Variables”, IEEE Transactions on Transactions on Intelligent Transportation Systems, 12 (3) pp. 695 - 704 (2011).
10. R. Lent, M. Minero, R. North, and J. Barria, “Evaluating Mobility Models in Participatory Sensing”, Proceedings of the first ACM international workshop on Mission-oriented wireless sensor networking (In conjunction with 18th ACM MobiCom 2012) pp. 3-8, Istanbul, August 22-26, 2012
11. S. Thajchayapong and J. A. Barria, "Anomaly Detection using Microscopic Traffic Variables on Freeway Segments," CD-ROM, Paper No. 10-2393, Transportation Research Board of the National Academies, Washington, D.C., 2010.
12. Q. R. Hamid and J. A. Barria, “Congestion Avoidance for Recharging Electric Vehicles using Smoothed Particle Hydrodynamics”, IEEE Transaction on Power Systems, Accepted for publication (2015).
13. Q. R. Hamid and J. A. Barria, “Distributed Recharging Rate Control for Energy Demand Management of Electric Vehicles”, IEEE Transaction on Power Systems, 28 (3) pp. 2688 - 2699 (2013).

7. More Information

Javier A. Barria
Intelligent Systems and Networks
Department of Electrical and Electronic Engineering
Imperial College London

http://www.imperial.ac.uk/people/j.barria

Email: j.barria@imperial.ac.uk

Phone: +44 (0)20 7594 - 6275
Fax: +44 (0)20 7594 - 6274