Essence

EU/H2020 GNFUV

Quality-aware Aggregation & Predictive Analytics at the Edge

NATASCHA HARTH & CHRISTOS ANAGNOSTOPOULOS

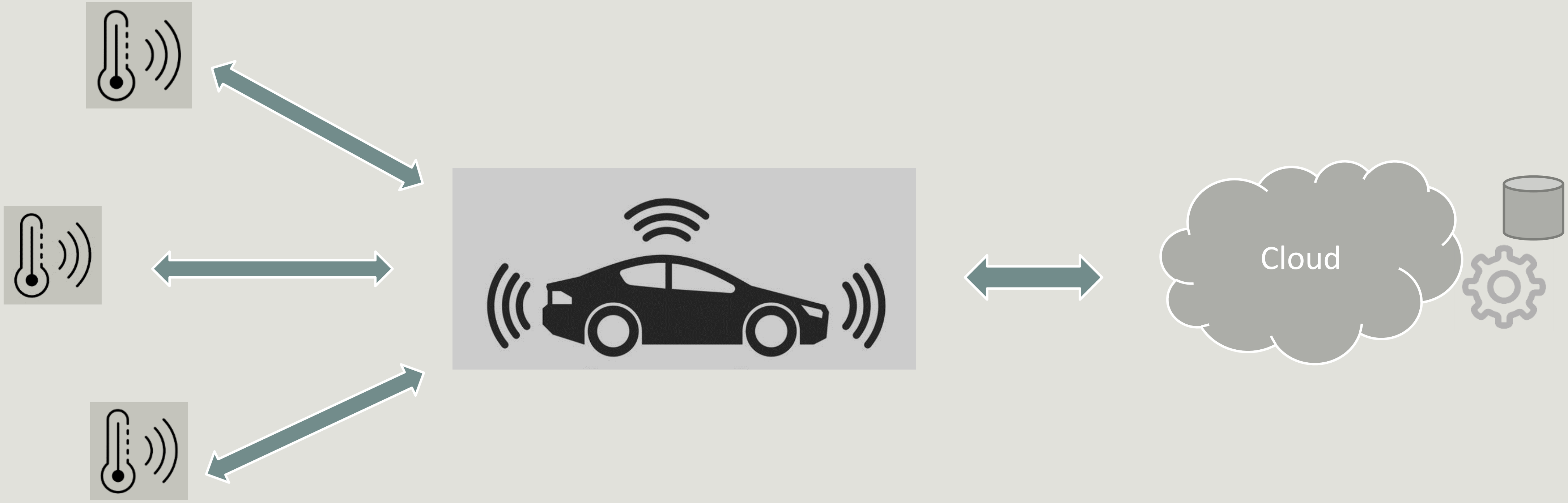
SCHOOL OF COMPUTING SCIENCE
UNIVERSITY OF GLASGOW, SCOTLAND

IEEE International Conference on Big Data ; 11-14 DEC 2017, Boston, MA, USA

Agenda

- Introduction
- Methodology
- Experiments
- Evaluation
- Results

Context



Sensing & Actuator Devices

IoT Gateways (Edge/Fog Network)

Cloud Environments

Constraints at the Edge

1. Limited Bandwidth…
2. Energy
3. Limited Computational Power
4. Storage Capacity
5. Latency!

Idea: Observe your Power & Push

- **Exploit the limited** computational power of sensing & actuator devices
- **Push Intelligence** to the Edge:
  - inferential tasks, on-line statistical learning, classification, localized detection,…**are pushed at the Edge**

Hypotheses & Actions

Given the **constraints** of an IoT network, let us **hypothesise** the following actions:

- **Action 1: Reduce** the communication overhead
  - **Hypothesis 1**: **not all data** are needed for inferential tasks/regression, i.e., **Learn More With Less**!
- **Action 2: Perform** real-time predictive analytics for instant action & autonomous decision making
  - **Hypothesis 2:** use the limited computational power to **infer** and take decisions in an **On-Line Manner**!
- **Action 3: Provide** high quality predictive analytics tasks (e.g., prediction accuracy, model fitting)
  - **Hypothesis 3:** decide **which** is the best data to learn and **when** to learn, i.e., **Be Intelligent On What You See**!

Challenges & Problem Definition

- Decide **which** data to communicate **without** losing quality of data & analytics
- **Problem 1:** time-optimized data selection problem.

- Decide **when** to deliver/send data and **what** to send in light of maximizing the predictive analytics accuracy
- **Problem 2:** time-optimized delivery scheduling problem.

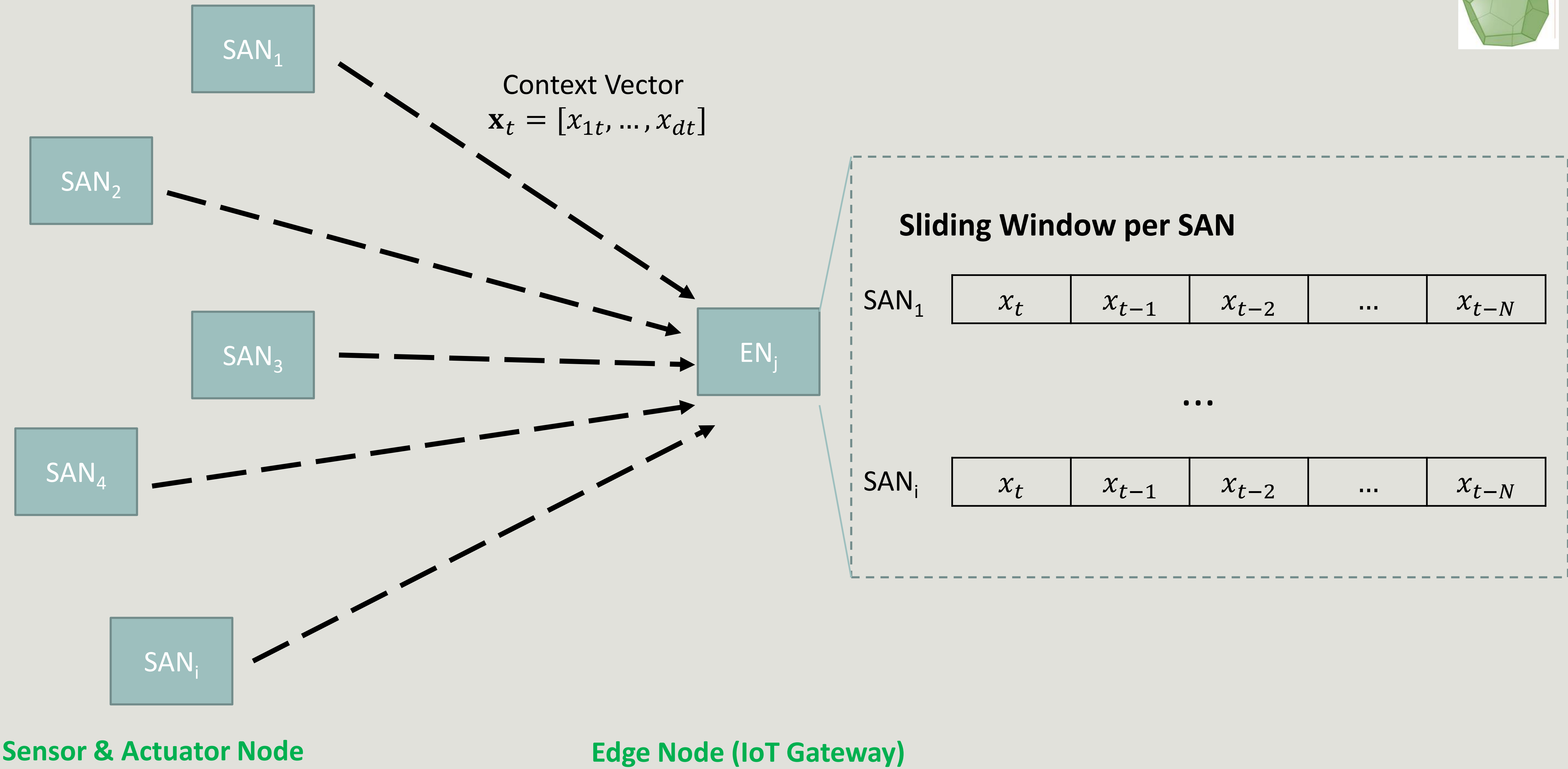
- Reduce **unnecessary** communication between/among devices and/or the Cloud
- **Problem 3:** conditionally data forwarding problem.

Contribution

- ✓Introduce an **optimal**, **quality-aware, and on-line** decision making model determining **when** and **which** data to deliver within the Edge Network.

- ✓Maximize the **quality** of analytics tasks s.t. being **communication efficient.**

- ✓**Domain:** Aggregation & Linear Regression Analytics over Sliding-Window Contextual Data Streams.



SAN$_1$
SAN$_2$
SAN$_3$
SAN$_4$
SAN$_i$
Context Vector
$\mathbf{x}_t = [x_{1t}, \dots, x_{dt}]$
EN$_j$
Sliding Window per SAN
SAN$_1$
x_t
x_{t-1}
x_{t-2}
…
x_{t-N}
…
SAN$_i$
x_t
x_{t-1}
x_{t-2}
…
x_{t-N}
Sensor & Actuator Node
Edge Node (IoT Gateway)

Idea: Predict, Decide & Reconstruct

Step 1: Local Prediction at SAN (Sensor & Actuator Node)

$$\hat{x}_t = f(x_{t-1}, \dots, x_{t-N}) = f_i(\mathcal{W})$$

$$e_t = d^{-\frac{1}{2}} \|x_t - \hat{x}_t\|$$

Step 2: Local Re-Construction at EN (Edge Node)

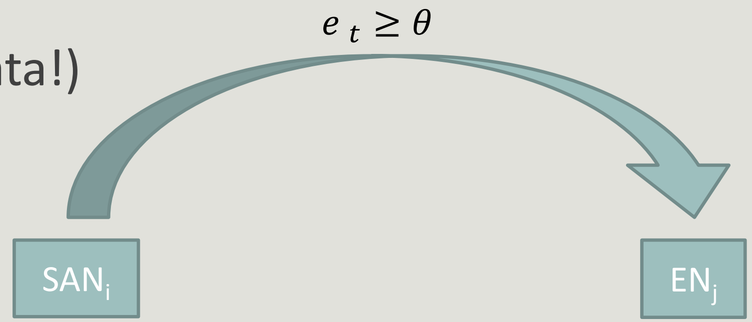
$$\tilde{x}_t = g(u_{t-1}, \dots, u_{t-M}) = g_j(\mathcal{W})$$

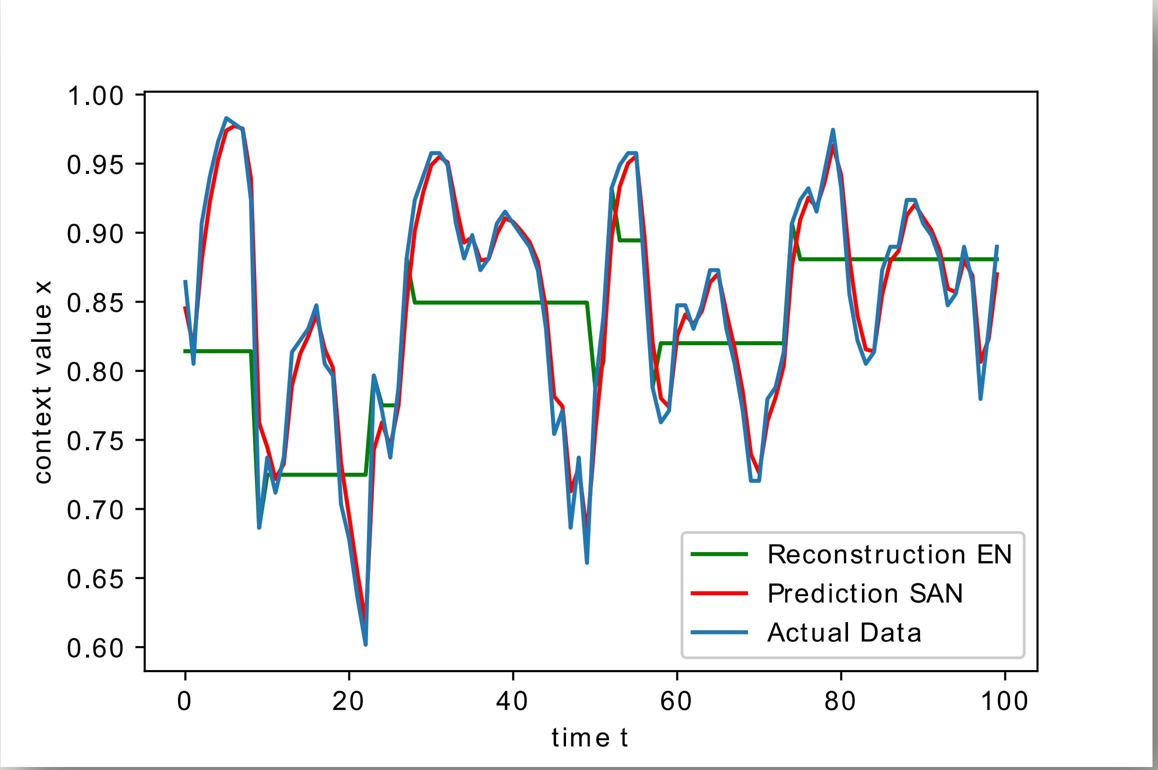
Instantaneous Decision Making (IDM)

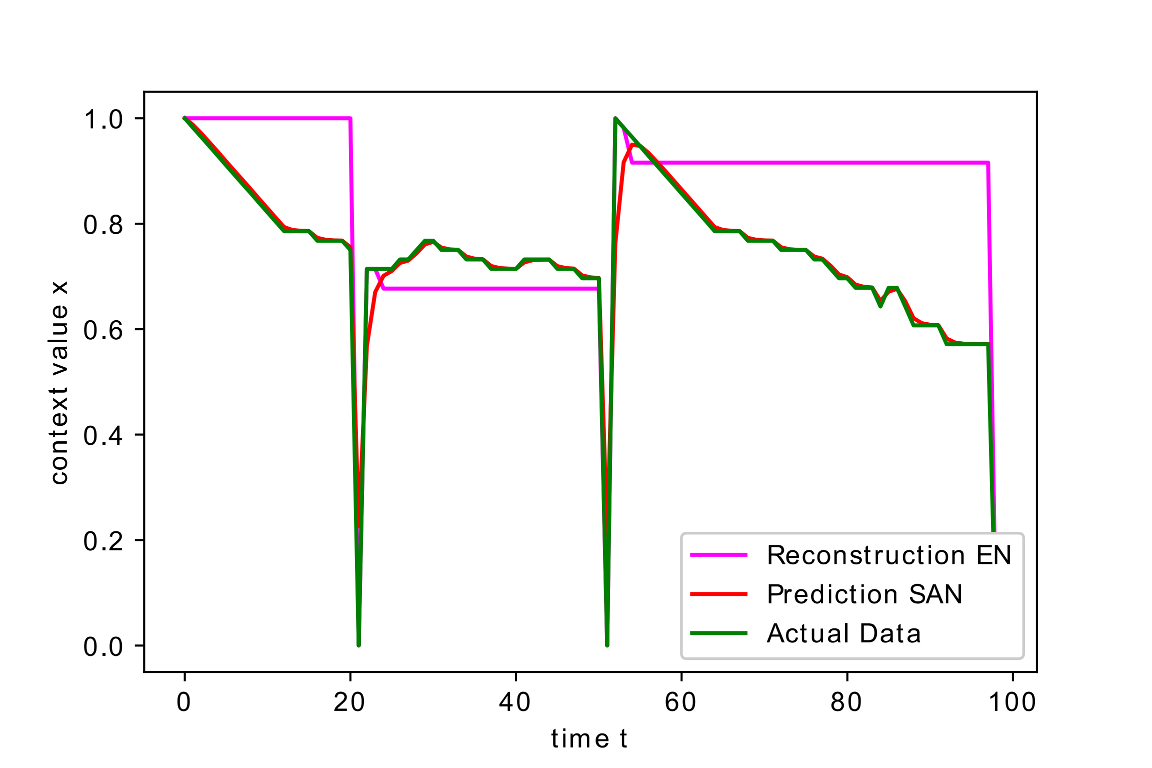
SAN employs locally selective forwarding: deliver data if **current** prediction error > threshold (ϑ)

- **Naïve Goal:** Reduce communication overhead; but no focus on the **Quality of Analytics**
- **Major Issues:**
  1. **What if** ϑ is relatively high (no control on the analytics quality)
  2. **What if** the prediction function in SAN is too good (never sends data!)
  3. **What if** Outliers occur (sends only outliers!)

  → information loss at the EN.







Observation: Very good prediction at SAN ☺
Consequence: EN cannot reconstruct data stream ☹

Observation: Outliers/Novelty data at SAN!
Consequence: EN receives only novelty/outliers ☹

Optimal Stopping Theory (**Which** & **When**)

Problem: IDM is not capturing the variability of the data stream inside EN…

Major Goal: Send **high** quality of information in the **right** moment, in other words: **which** data and **when** to send.

Solution: Develop an Optimal Stopping Time stochastic model to find the **optimal forwarding time** at SAN such that **maximises** the expected analytics quality at the EN.

Idea: Instead of sending every time θ exceeds prediction error, we find the **best time** and **context vector** to send:

- **we optimally delay** data delivery thus being communication efficient;
- **we accumulate** the prediction error history of IDM decisions thus controlling the analytics accuracy.

Optimal Vector Forwarding (OVF)

Induced delay is based on the history of prediction error (Quality Tolerance):

$$Z_t = \begin{cases} \lambda\theta & \text{if } e_t > \theta, \\ e_t & \text{if } e_t \leq \theta. \end{cases}$$

Monitor: Accumulation of Local Prediction Errors

Quality Tolerance Reward:

$$Y_t = \beta^t S_t = \beta^t \sum_{\tau=0}^{t} Z_\tau$$

Tolerate: Do not send & accumulate prediction errors

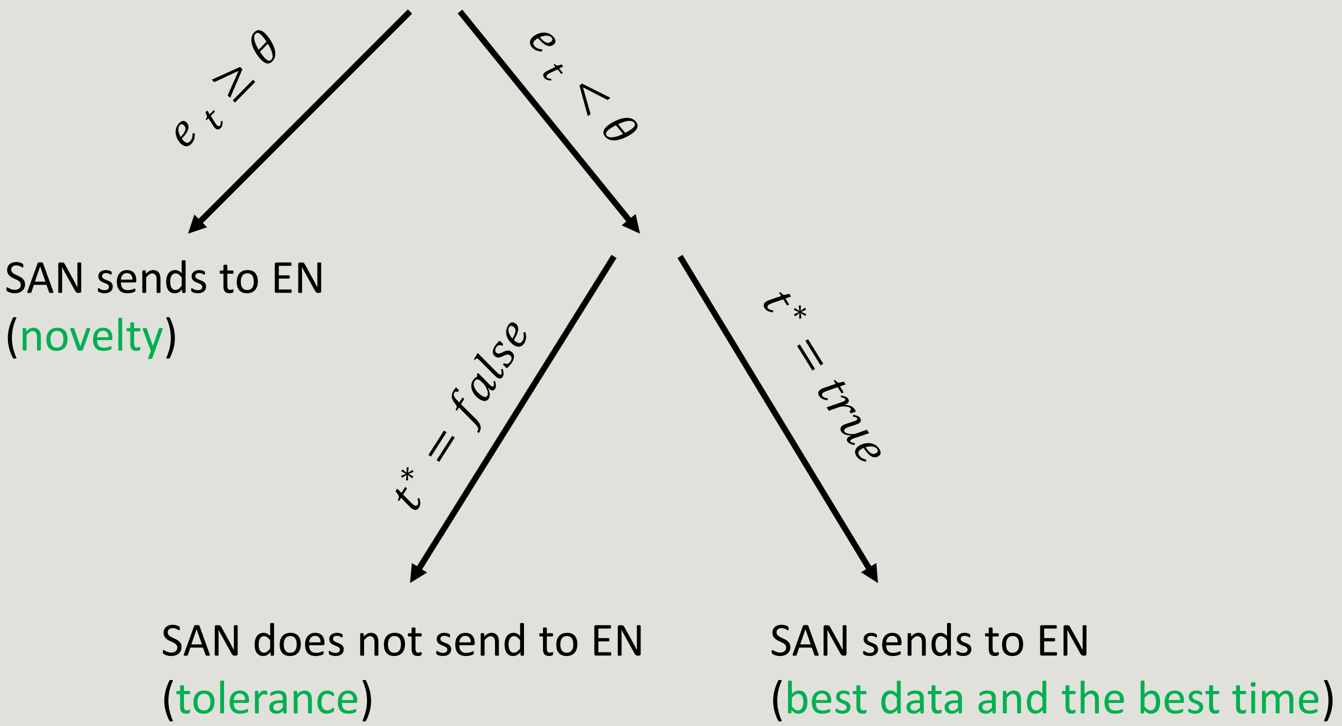
Optimal Stopping Time: We send data from SAN to EN when...

$$t^* = \inf\left\{t \geq 1 \,\middle|\, \sum_{k=1}^{t} Z_k \geq \frac{\beta}{1-\beta} E[Z]\right\}$$

Minimize communication overhead
s.t. maximizing quality tolerance

Hybrid Optimal Vector Forwarding (HOVF)

Combines IDM and OVF methodology



Methodologies

1. **Baseline**: Sending continuously data from Edge to Cloud!

2. **Instantaneous Decision Making** (IDM): prediction-error based decision

3. **Optimal Vector Forwarding** (OVF): quality-tolerance based decision

4. **Hybrid Optimal Vector Forwarding** (HOVF): intelligence is now pushed at SANs

Experiments

Exponential Smoothing for prediction (SAN) and reconstruction (EN)

$$s_t = \alpha x_t + (1 - \alpha) s_{t-1}$$

Analytics tasks at the EN:

- Aggregation Analytics (sliding window-based AVG, MEDIAN, MAX, ...)
- Multivariate Linear Regression (sliding window-based regression)

Real datasets:

- Air Quality 4-dim contextual vectors;
- Environmental 4-dim contextual vectors in the School of Computing Science, Uni of Glasgow.

Evaluation: Three Directions

1. **Communication:** number of messages sent from SANs to EN (communication overhead);
2. **Information:** quality of data at the EN (information theoretic perspective);
3. **Analytics quality at the EN:**
   a) Re-construction error w.r.t. ground truth;
   $$\alpha_t = \|x_t - \tilde{x}_t\|$$
   b) Aggregation analytics discrepancy w.r.t. ground truth;
   $$\gamma = \|h(W) - h(W^*)\|$$
   c) Regression performance discrepancy w.r.t. ground truth;
   $$\delta = \|\epsilon - \epsilon^*\|$$
   d) Model Fitting discrepancy w.r.t ground truth;
   $$\delta' = \|\boldsymbol{w} - \boldsymbol{w}^*\|$$

Metrics

Discrepancy is evaluated w.r.t. baseline solution (sending all data)

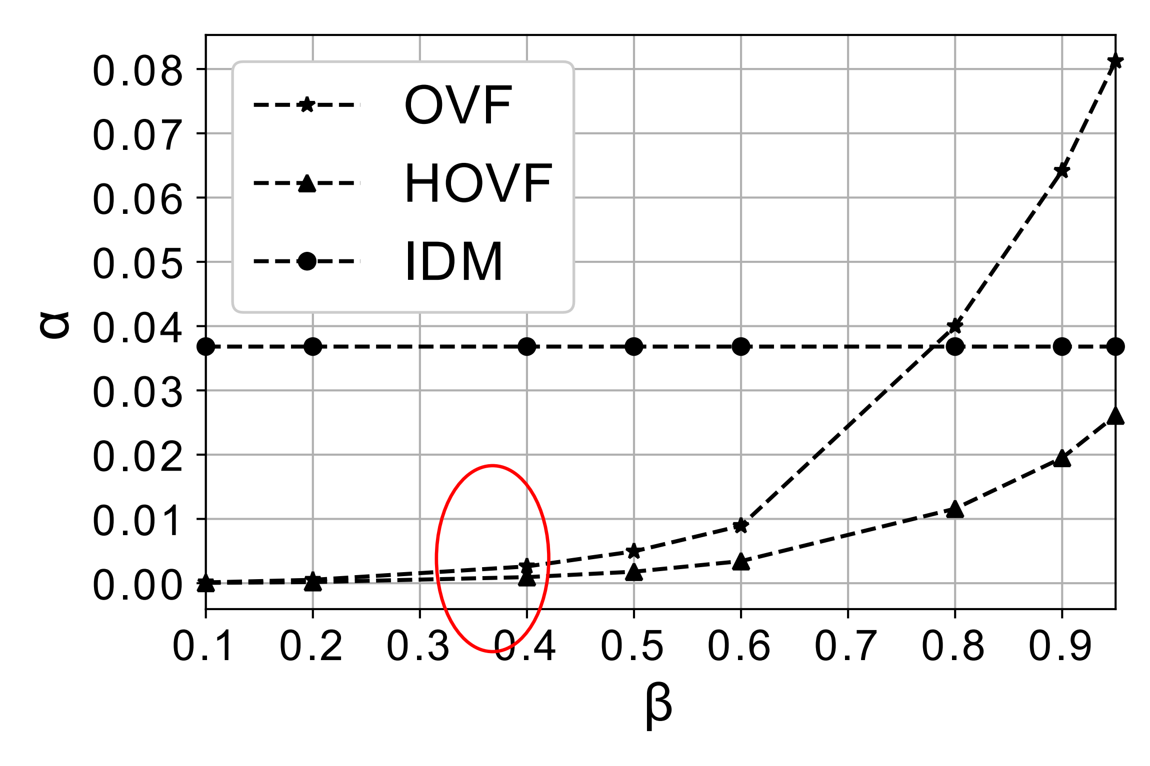
1. **Communication**
   - Percentage of remaining communication w.r.t the baseline solution

2. **Analytics quality**
   - For α and γ → Symmetric Mean Average Percentage (SMAPE)
   - For δ →Root Mean Squared Error (RMSE)

3. **Information loss**
   - Kullback-Leibler (KL) divergence

$$KL(p(\tilde{x}) \parallel p(x)) \ \int_0^1 p(\tilde{x}) \log \frac{p(\tilde{x})}{p(x)} \mathrm{dx}$$

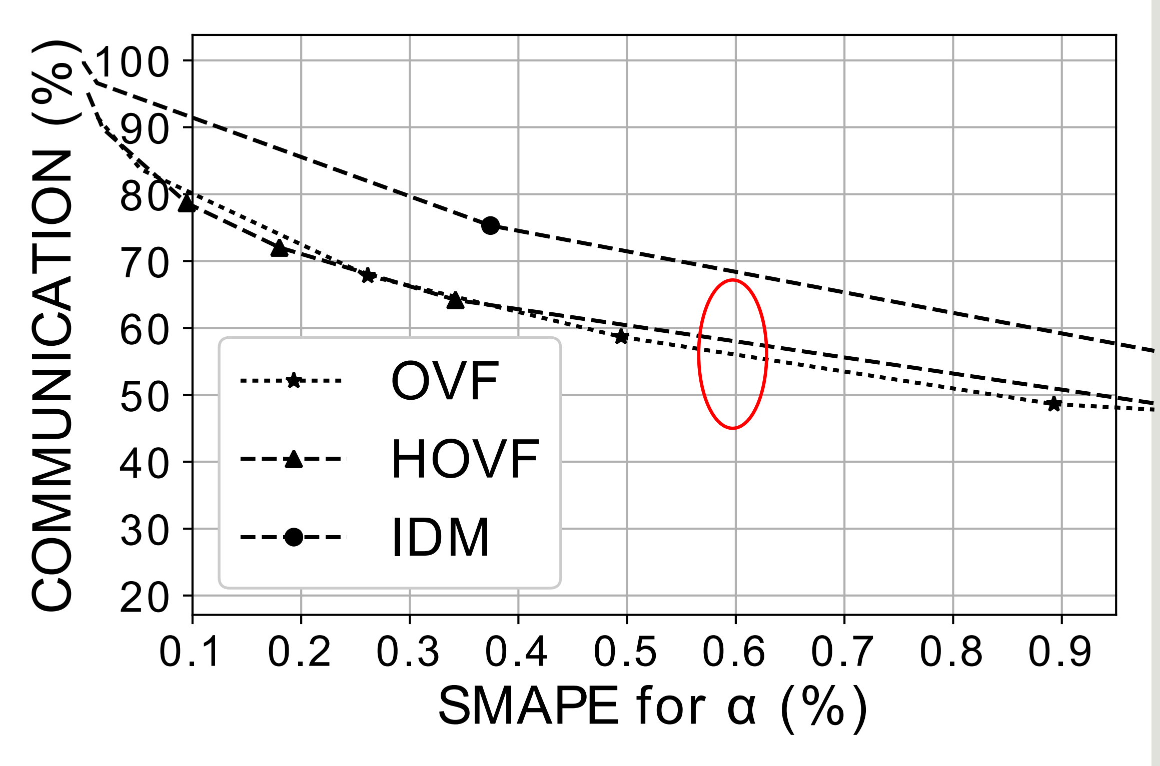
Results



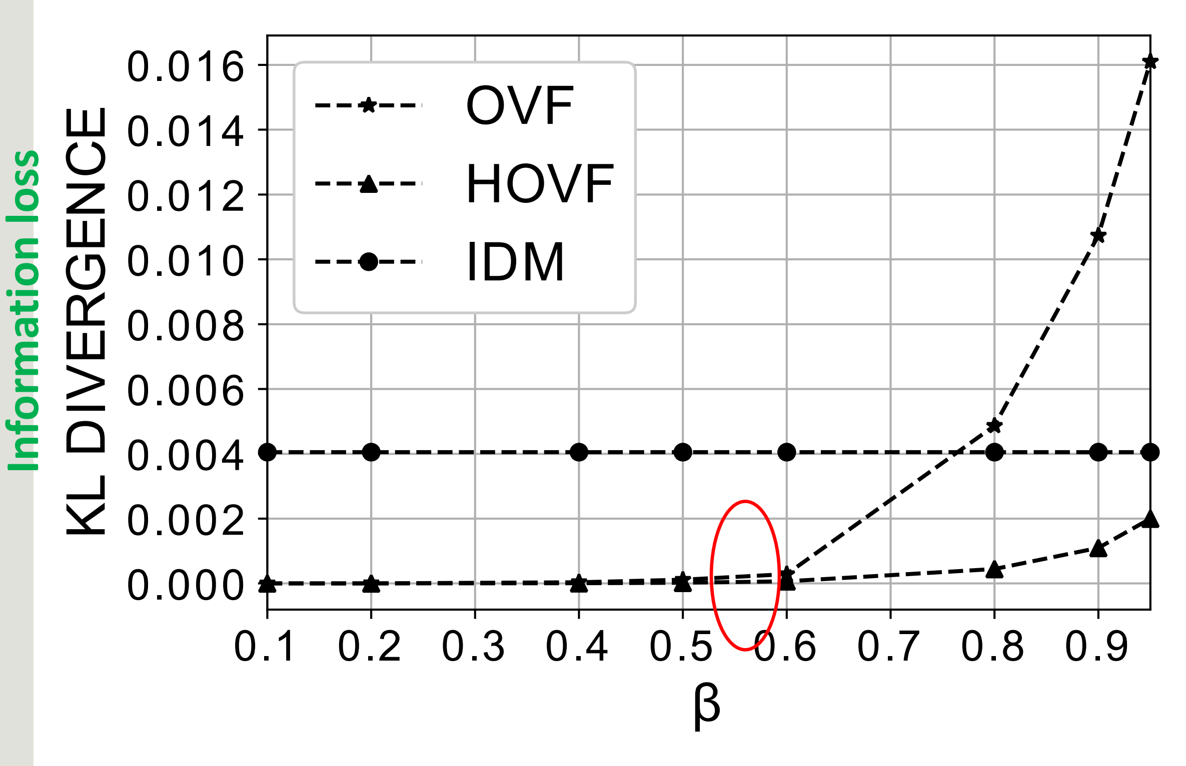
reconstruction

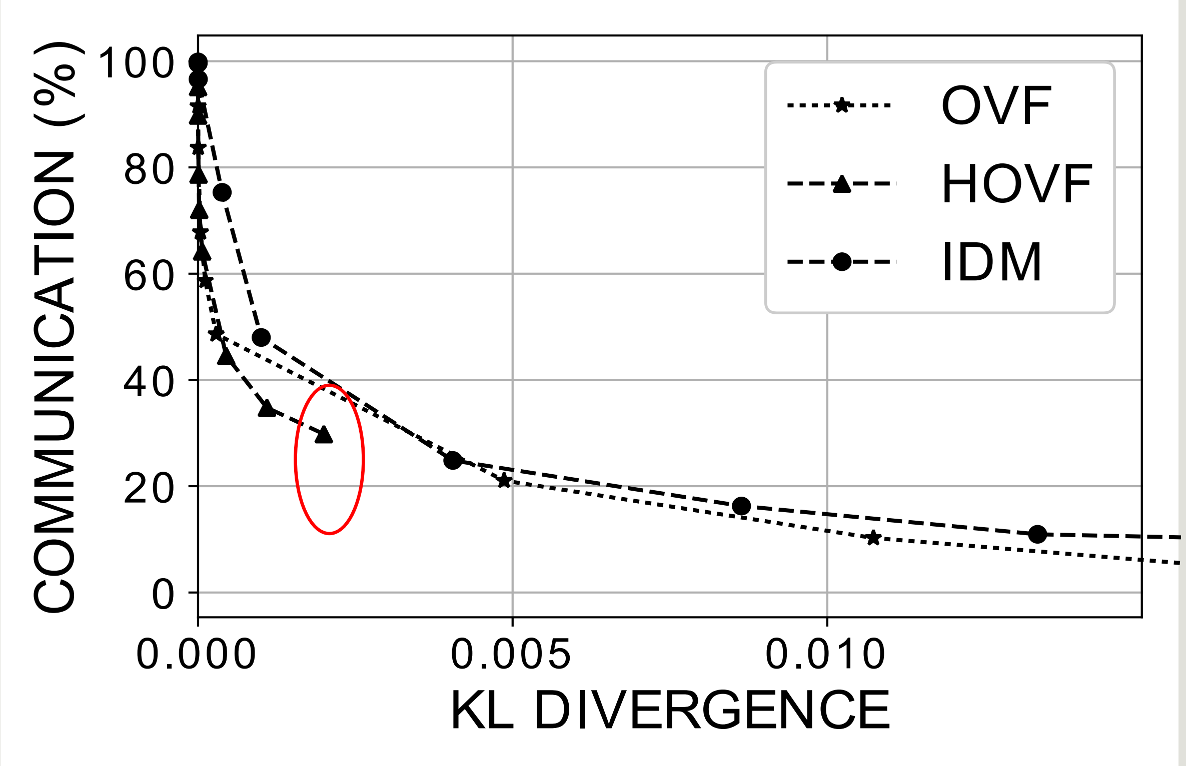
tolerance factor

Efficiency (quality vs communication)



Efficiency (quality vs communication)





tolerance factor

University of Glasgow

Essence

EU/H2020 GNFUV

THANK YOU!

Natascha Harth: n.harth.1@research.gla.ac.uk
Christos Anagnostopoulos: Christos.Anagnostopoulos@glasgow.ac.uk

Essence: Pervasive & Distributed Intelligence: http://www.dcs.gla.ac.uk/essence/