QUEEN'S UNIVERSITY BELFAST

ESTD 1845

QUEEN'S UNIVERSITY BELFAST

DISTRIBUTED STREAM PROCESSING ON EDGE AND FOG COMPUTING VIA TRANSPRECISE COMPUTING

HANS VANDIERENDONCK

DiPET COORDINATOR

28 MARCH 2022

SCIENTIFIC BACKGROUND (1) - MOTIVATION

Edge and fog computing
Distributed computing paradigms focused on performing services and computations closer to end users. SDCDN considers the question of distributing computations across networks.

Stream processing
Stream processing encompasses a class of applications that operate on periodic events or data streams in a dataflow-like fashion.

Transprecise computing
Traditionally, computing has been designed to be exact, often resulting in excessive computation and data footprints. Transprecision aims to adapt precision as requirements and compute power change.

Quality of Service / Experience
Edge/fog performance quantified through QoS/QoE metrics; align closely to precision (detail in computation) and accuracy (value of computation).

QUEEN'S UNIVERSITY BELFAST

SCIENTIFIC BACKGROUND (2) - SCENARIOS

Malleable applications
Adapt resource usage and QoE of streaming applications at runtime based on availability of resources and/or characteristics of the data.

Deploying applications on insufficient resources
Where applications require more compute power are memory than a device can provide, it may be preferable to execute it at reduced QoE rather than not at all.

Migration of applications
Availability of resources can be a trigger for application migration, as well as a contributing factor to decide which applications to migrate and where to migrate them to.

Oversubscription of resources
Quantifiable loss of QoE in the presence of temporary oversubscription of resources. Instead of applications competing for resources, QoE-tolerant applications can be scaled back.

QUEEN'S UNIVERSITY BELFAST

PROJECT STRUCTURE

WP1 Streaming algorithms with tunable precision (UVT)
New algorithms where precision can be controlled

WP2 Programming model and runtime implementation (FORTH)
Extending Apache Flink to express QoE and tunable precision parameters in the programs

WP3 Dynamic resource mapping and middleware (UR1)
Extending Kubernetes resource allocation, scheduling and scaling for transprecision

WP4 Deployment and evaluation (UPC)
Evaluate and demonstrate efficacy of key use cases in a realistic network environment

QUEEN'S UNIVERSITY BELFAST

KEY CHALLENGES AND POTENTIAL IMPACTS

CHALLENGES

Characterising precision of applications and designing autonomous systems to do so

Controlling the execution flow of applications through transprecise parameters

Operating without run-time feedback on the quality of choices

IMPACT REALISATION

Collaboration with the **FlockFocus** project at QUB to deliver real-time video analytics

- Tracking and behaviour analytics of chicken flocks at chicken farms
- Co-sponsored by Foundation for Food and Agriculture Research and McDonald’s

Collaboration between Inria and **OVHcloud**

- Reducing energy consumption of cloud data centers
- Exploit automatic scaling methods for transprecision-aware data stream processing

PARTNERS

Coordinator, Lead WP5 Management
Transprecise computing (Lee, Vandierendonck)
Edge computing and middleware (Varghese)

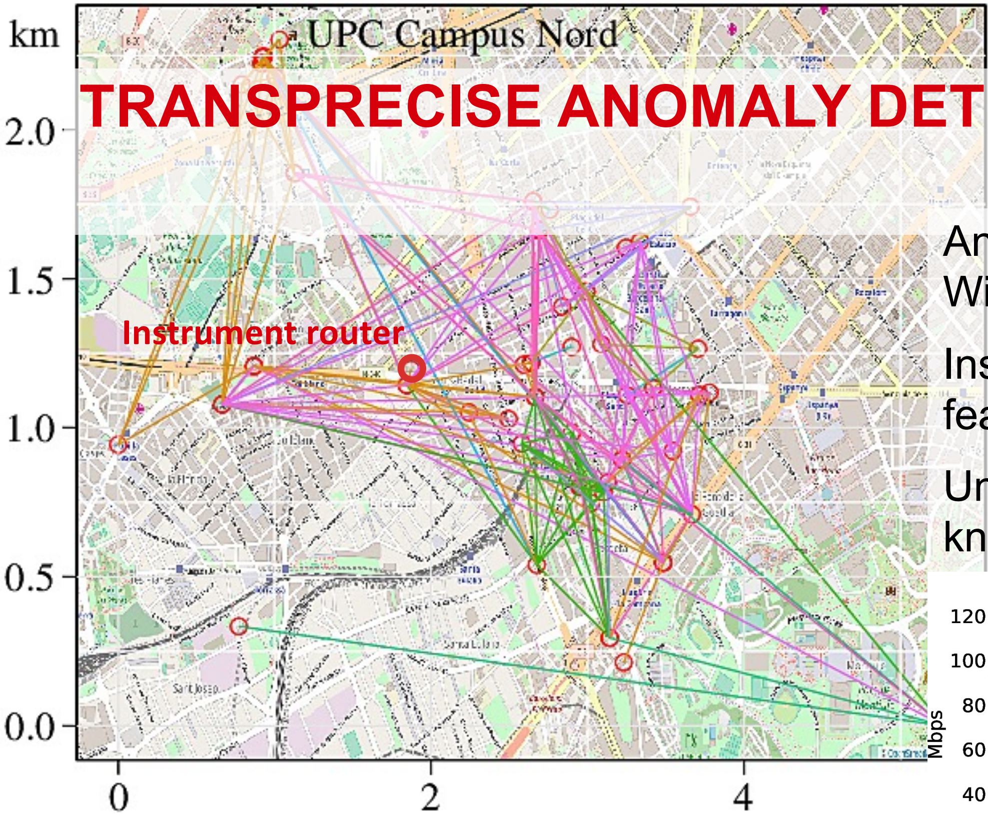
Lead WP2 Programming Model and Runtime Implementation
Programming languages and runtimes, analytics (Pratikakis)

Lead WP3 Dynamic Resource Mapping and Middleware
Fog computing (Pagliari, Mondal, Pierre)
Geo-distributed data stream processing (Pagliari, Mondal, Pierre)

Lead WP4 Deployment and Evaluation, community network testbed
Networked distributed systems (Vilchez, Freitag, Navarro)
Wireless Community Networks (Cerdà-Alabern, Baig)

Lead WP1 Streaming Algorithms with Tunable Precision
AI/ML, Distributed Computing (Iuhasz, Petcu)
HPC, Cloud Computing, Service Oriented Computing (Panica)

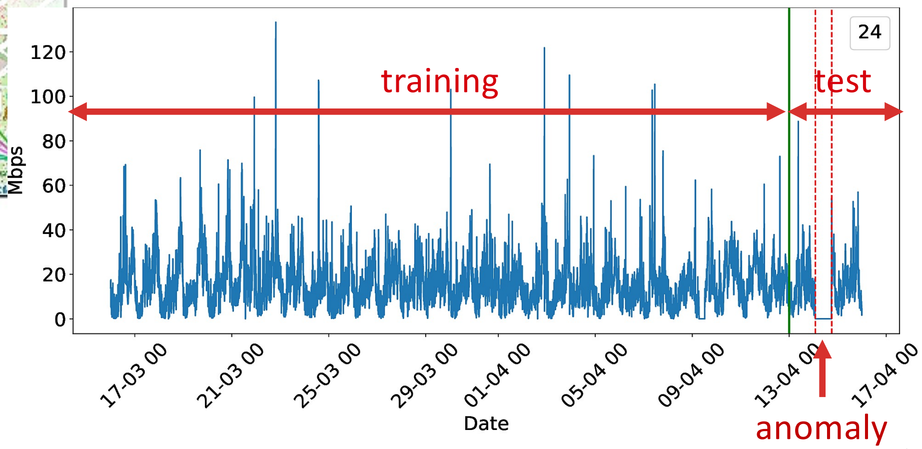
TRANSPRECISE ANOMALY DETECTION



Anomaly detection in a production Wireless Community Network

Instrument router to collect traffic stats: 2397 features, 7931 samples 5 minutes apart

Unsupervised machine learning algorithms detect a known anomaly (router outage)



Traffic feature of a node showing the training and testing sets (left and right of the Green line) and the failure interval (red lines).

Dissemination

Collaboration between UVT and UPC

Open Dataset DOI: 10.5281/zenodo.6169917

Conference paper presented, September 2021

Journal paper in preparation

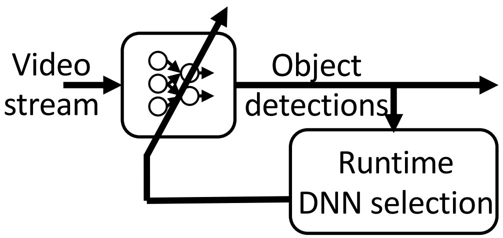
VIDEO ANALYTICS AT THE EDGE: TRANSPRECISE OBJECT DETECTION

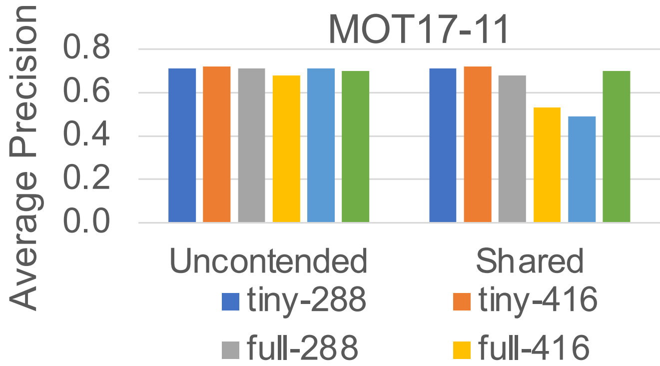
DiPET

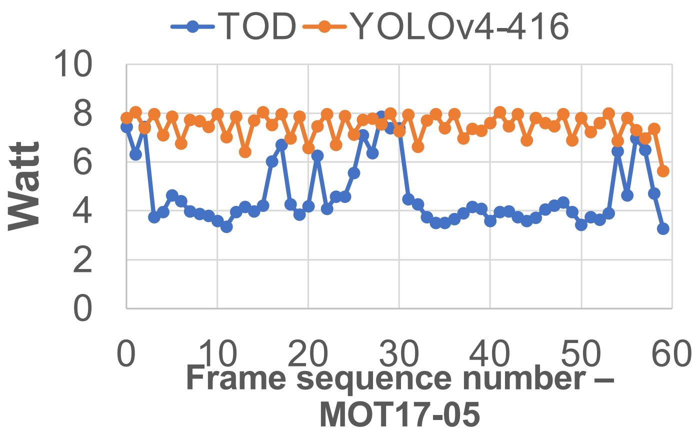
Resource constraints on edge computing devices require careful application of deep neural networks

Trade-off between network complexity, object size, speed of movement, and frame rate

Runtime decision-making to select 1-of-N YOLOv4 neural networks, dependent on video content







Lee, Varghese, Woods, Vandierendonck, "*TOD: Transprecise Object Detection at the Edge*", IEEE ICFEC, 2021

SUSTAINABILITY AND VALORISATION

INDUSTRY ENGAGEMENT

- QUB through the ECIT engineering team (knowledge transfer support mechanism) and its industry partnerships
- UPC will engage with eXo to apply transprecise network monitoring to support network management

FOLLOW-ON PROJECTS

- UR1 will leverage the autoscaling mechanisms to minimise reconfigurations (TERRA FORMA, France)
- UVT will continue the development of the event detection engine (EDE), previously designed for exascale systems, in follow-on projects
- QUB will continue the exploration of transprecise computing in future projects (EU MSCA IF Fellowship SoftNum)
- Additionally, the concepts are disseminated through undergraduate and graduate courses and research projects, as well as PhD projects, by all partners