

# A Node-Level Managing System for ML-based Autonomic Operation

Alexandros Patras (patras@uth.gr)

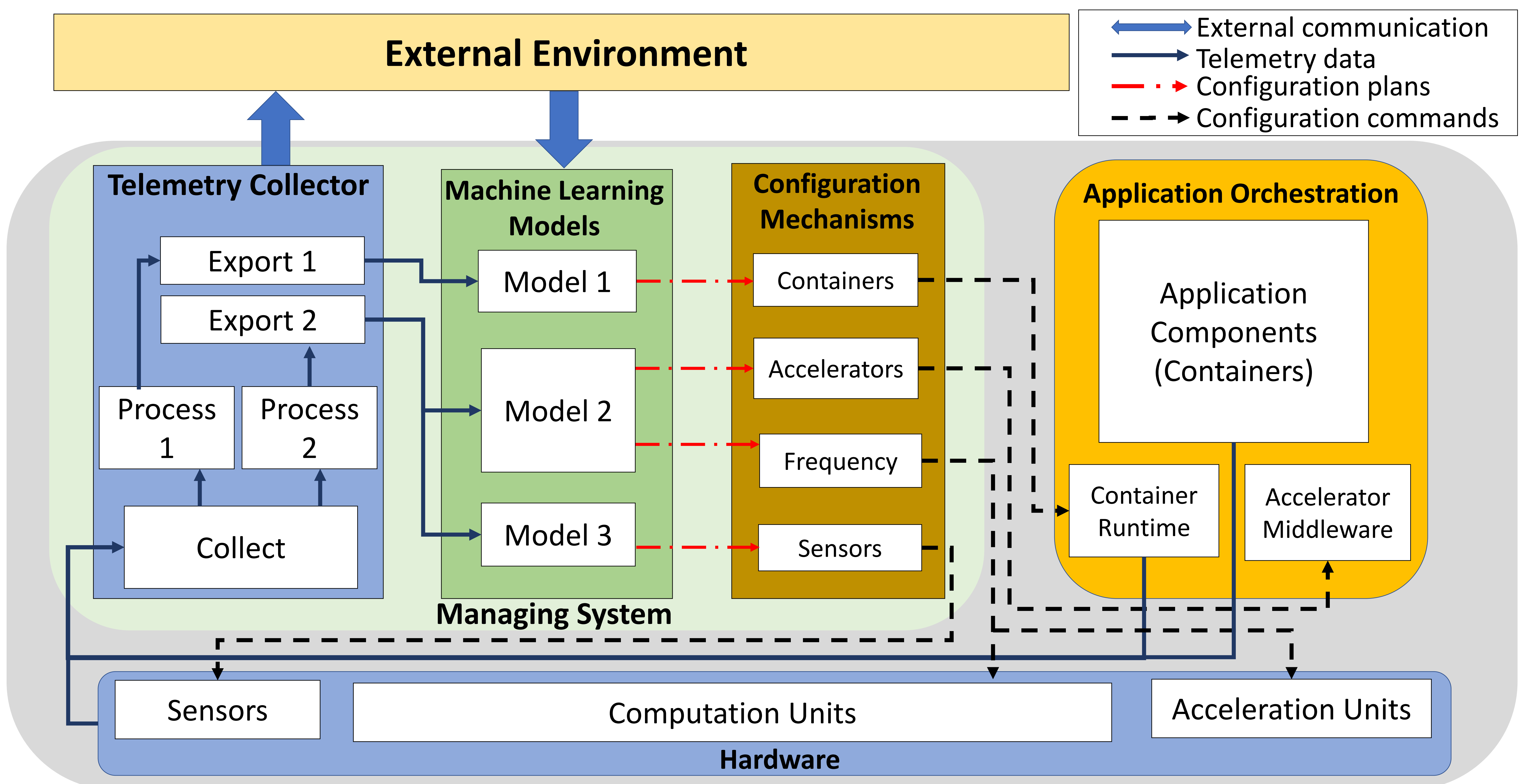
Foivos Pournaropoulos (spournar@uth.gr)

## Motivation

- Application containers may be deployed on a cluster of nodes that are heterogeneous and physically distributed.
- There are many configuration options available, spanning from which node the computation will be placed on to how the node will be configured.
- A human operator might not be able to cope with the complexity of this configuration span.

## Managing System

- There are 3 main entities:
  - I. Telemetry Collector
  - II. Machine Learning Models
  - III. Configuration Mechanisms
- Different machine learning models that produce configuration plans for different mechanisms.
- The telemetry system has been implemented using the OpenTelemetry specification and includes application metrics (e.g. QoS).



## Work in Progress

Developing an interface based on OpenAPI for available configuration mechanisms:

- CPU/GPU Frequency
- Container CPU Core Pinning
- Which processor is used (CPU/Accelerator)
- Sensing parameters

## Future Work

- Explore machine-learning models that will exploit the mechanisms to achieve optimal operation.
- Study the effects of external factors, that can affect the node-level operation.



## Acknowledgments

This work has been supported by the European Community's Horizon Europe Program under Grant Agreement nr. 101092912, project MLSysOps.

