HELLENIC REPUBLIC
UNIVERSITY OF THESSALY

Optimized Analytics Query Allocation at the Edge of the Network

Anna Karanika, Madalena Soula,
Christos Anagnostopoulos,
Kostas Kolomvatsos & George Stamoulis

The 12th International Conference on Internet and Distributed Computing Systems
Napoli, Italy
10-12 October 2019

Outline

1 **Introduction**
Current state of the art, our contribution and novelty of our work

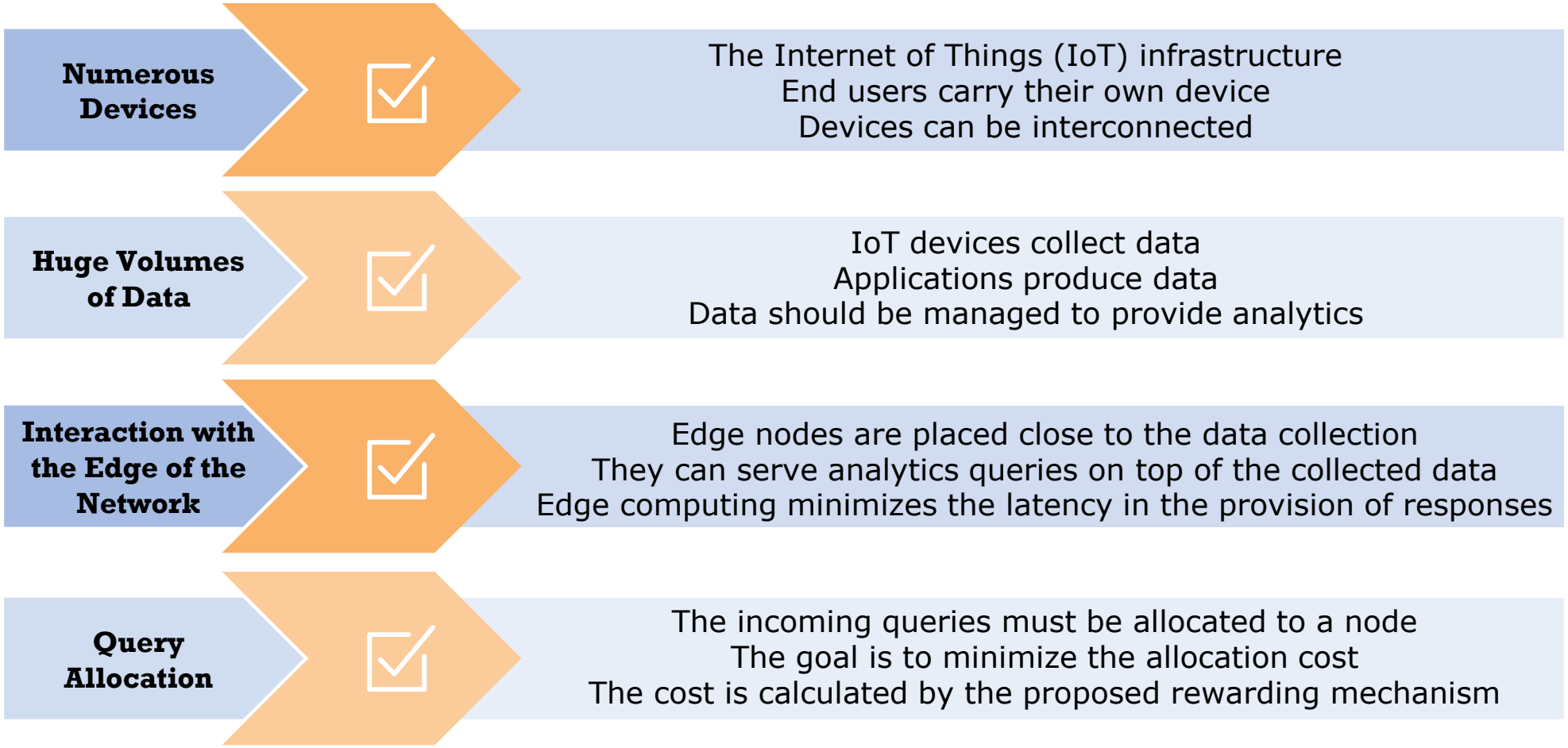
2 **Problem Description**
Description of the envisioned setting

3 **The Proposed Model**
Our approach for the construction of a rewarding mechanism

4 **Experimental Evaluation**
Description of our experiments and the delivered outcomes

5 **Conclusions**
Our conclusions and vision for future work

Introduction



Introduction

State of the Art

- Analytics models/systems mainly focus on reducing response time
- They target the minimization of data transfer in the network
- The methods adopted to this end are sampling and progressive analytics

Our focus

- We incorporate the dynamics of the environment where IoT devices act
- Our goal is to allocate incoming queries to nodes efficiently
- We treat our problem as an assignment problem

We propose

- We propose the use of a simple rewarding mechanism
- We take into consideration:
  - the query's complexity & deadline
  - the node's load & speed
- We adopt the Hungarian Algorithm as modified by Munkres

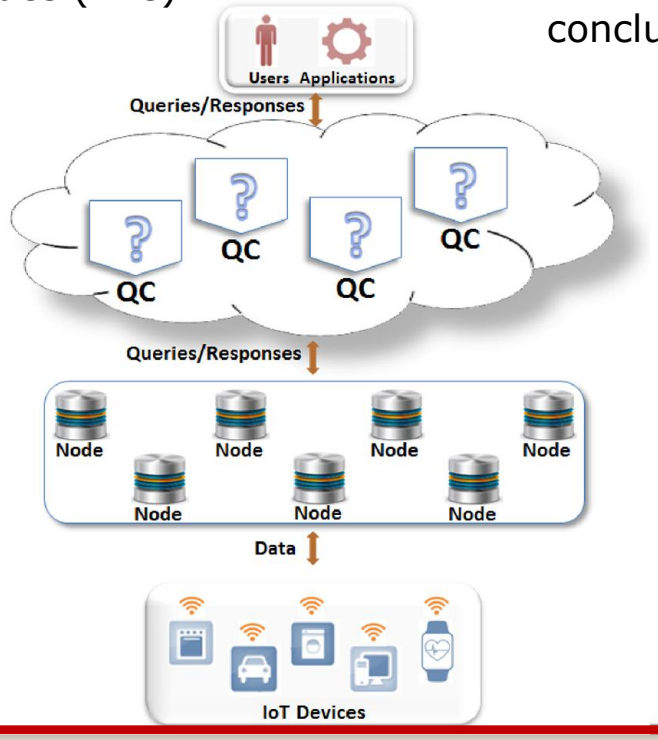
The Envisioned Setting

IoT devices collect multivariate data from their environment while users/applications generate queries

Query Controllers (QCs) orchestrate the allocation of the queries to the Edge Nodes (ENs)

The proposed mechanism indicates the cost that QCs should pay for every allocation they desire to conclude

We consider a batch-oriented approach as well as a sliding window approach



The Envisioned Setting

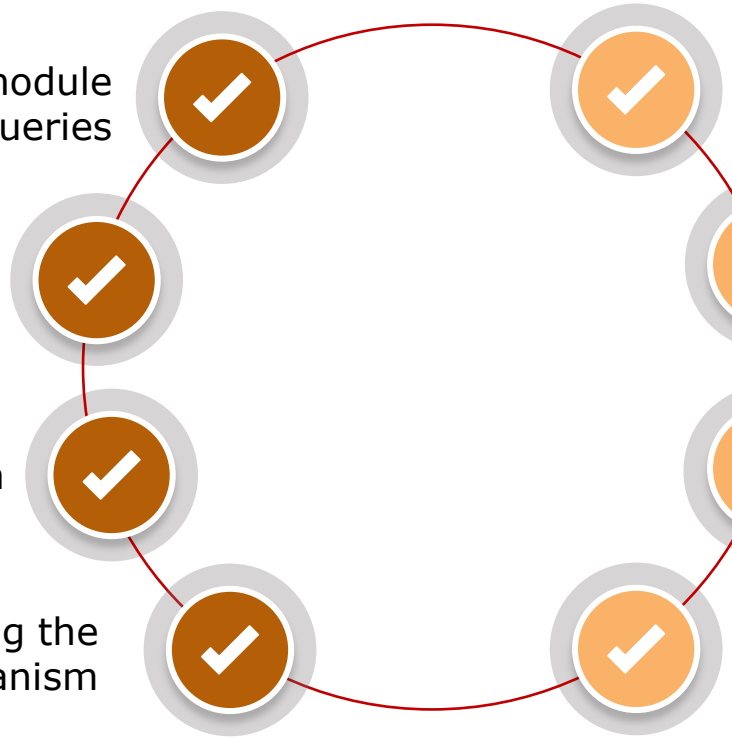
Query Controller

The Query Controller is a module which manages the incoming queries

It should efficiently respond in the minimum possible time

We adopt the Hungarian Algorithm

We create a cost matrix using the proposed rewarding mechanism



Queries–Edge Nodes

As an allocation, we define the optimal selection of a node for a distinct query

The allocation is determined by the query's characteristics and the node's current performance

Query's Characteristics:

i. complexity, c_{q_t}

ii. deadline, τ_{q_t}

Node's Characteristics:

i. load, l_j

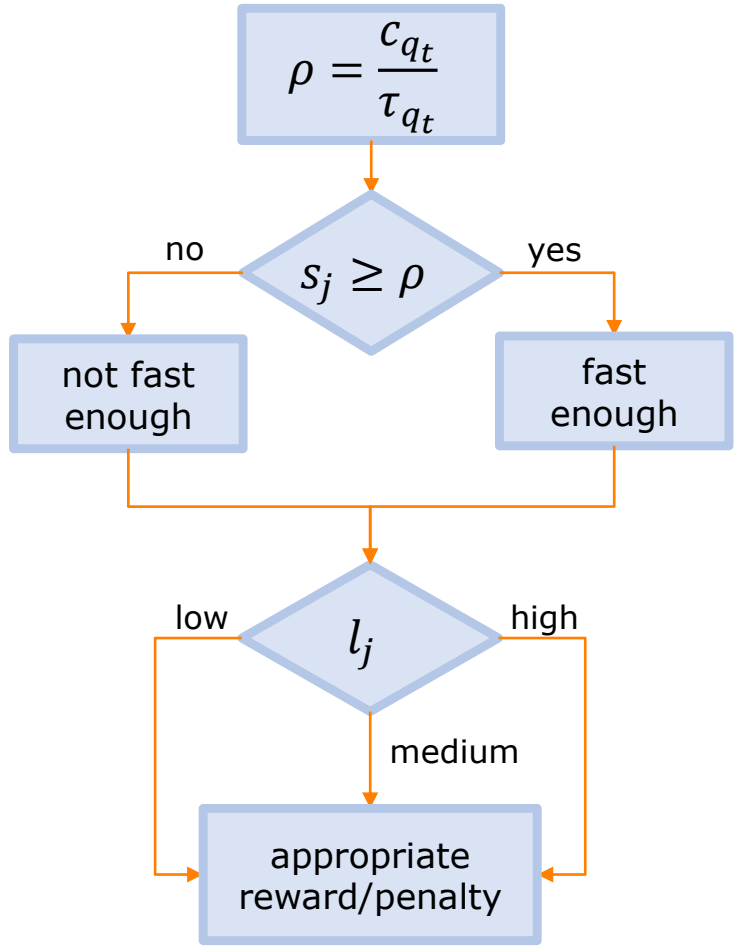
ii. speed, s_j

The Rewarding Mechanism

The rounded ratio of the query complexity compared to the query deadline is calculated, representing the speed demanded for the query's execution within its deadline.

The node's speed is compared to the ratio. Thus, the node is defined either fast enough to serve the query or not.

For each speed and load combination, an appropriate reward or penalty is attributed to the allocation cost.



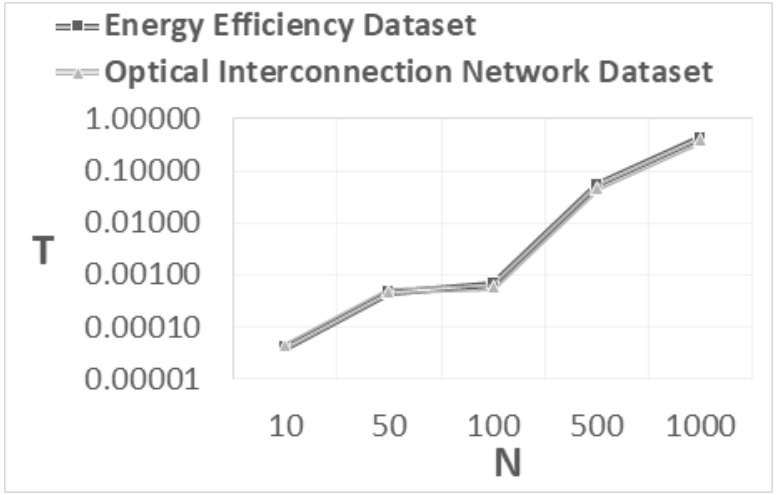
Experimental Setup

- We report on the performance of the model
- We aim to reveal the efficiency of the envisioned allocations
- We adopt several performance metrics to evaluate the results

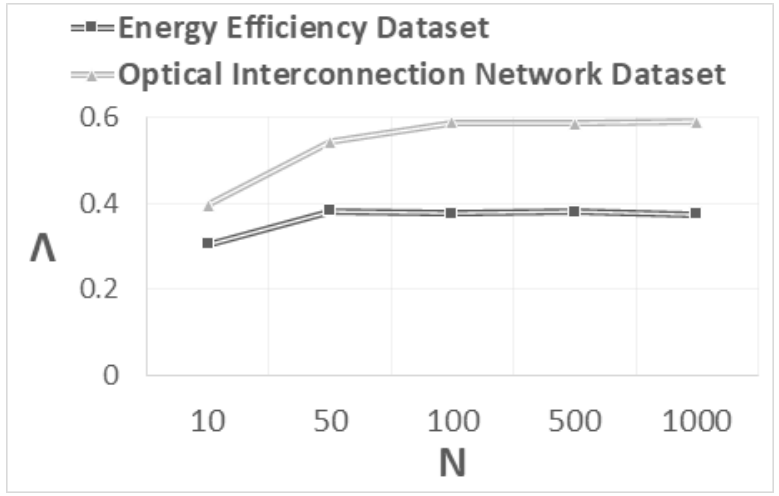
Metric	Description
T	the time required for concluding an allocation
Λ	the difference of the selected node's load with the lowest load among all nodes $(\Lambda = l_{selected} - l_{lowest})$
Σ	the difference of the highest speed among all nodes with the speed of the selected node $(\Sigma = s_{highest} - s_{selected})$
Φ	is a linear combination of the Λ and Σ metrics $(\Phi = \alpha * \Lambda + (1 - \alpha) * \Sigma,\ \alpha \in [0,1])$

Dataset	Source
Energy Efficiency Data Set	https://archive.ics.uci.edu/ml/datasets/Energy+Efficiency
Optical Interconnection Network Data Set	https://archive.ics.uci.edu/ml/datasets/Optical+Interconnection+Network+

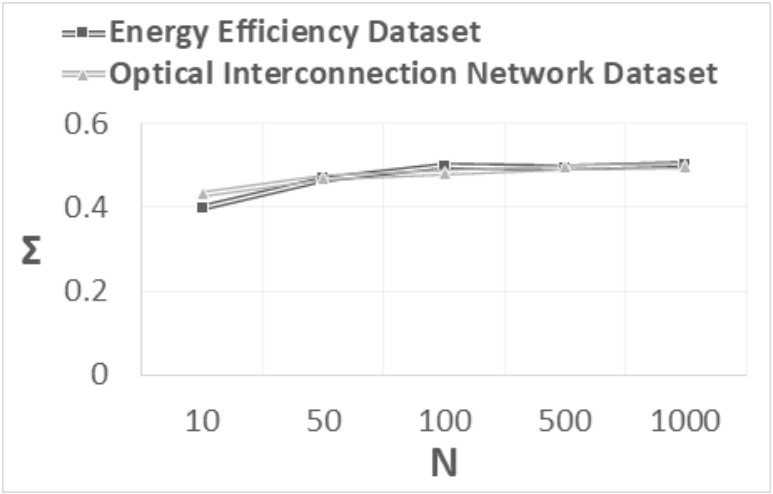
Experimental Results



$T \propto N$, T increases linearly

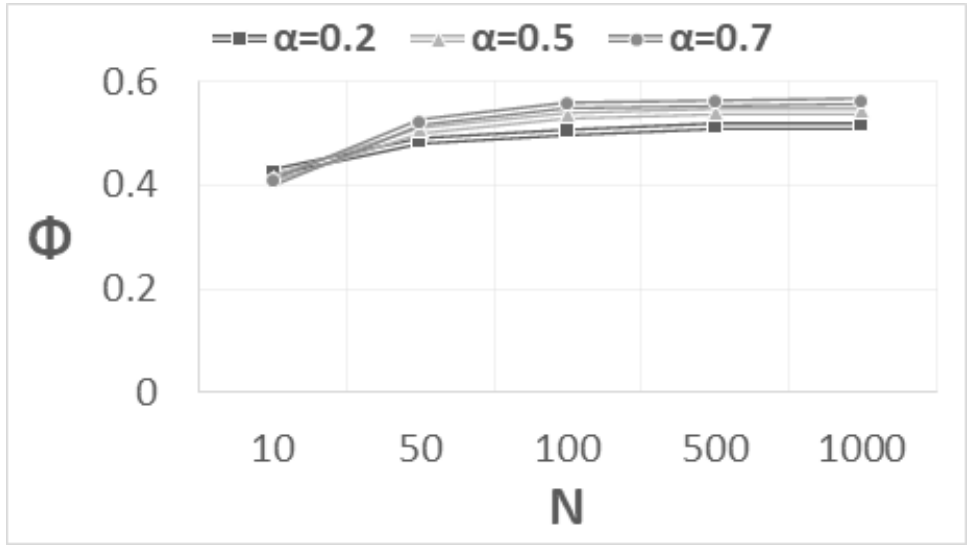


The average difference with lowest possible load is around 0.3-0.4

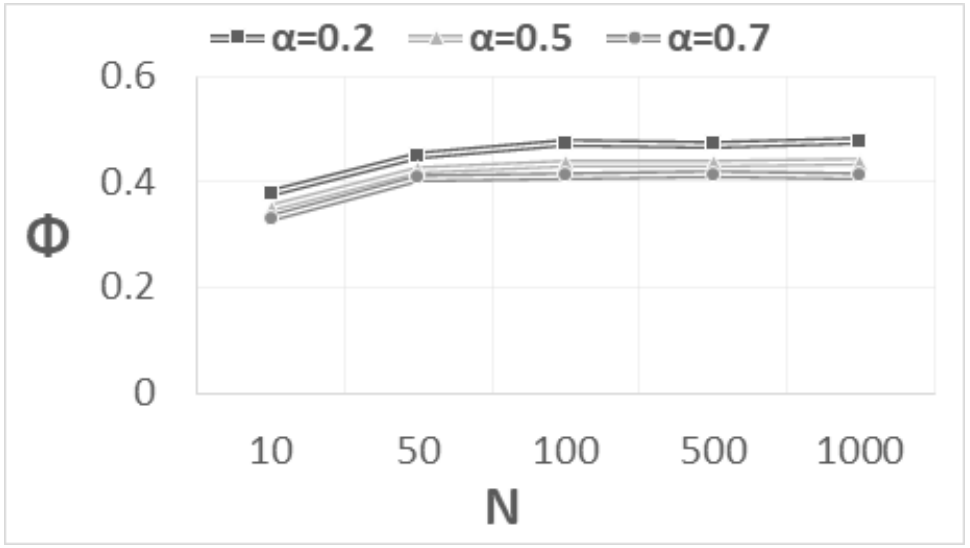


The average difference with the highest possible speed is around 0.5

Experimental Results



Results for Φ the metric – Optical Interconnection dataset



Results for Φ the metric – Energy Efficiency dataset

- $\alpha = 0.2$: low difference with the optimal node concerning the speed
- $\alpha \to 1.0$: low difference with the optimal node concerning the load

The proposed model exhibits a stability while trying to incorporate the optimal decision for both characteristics

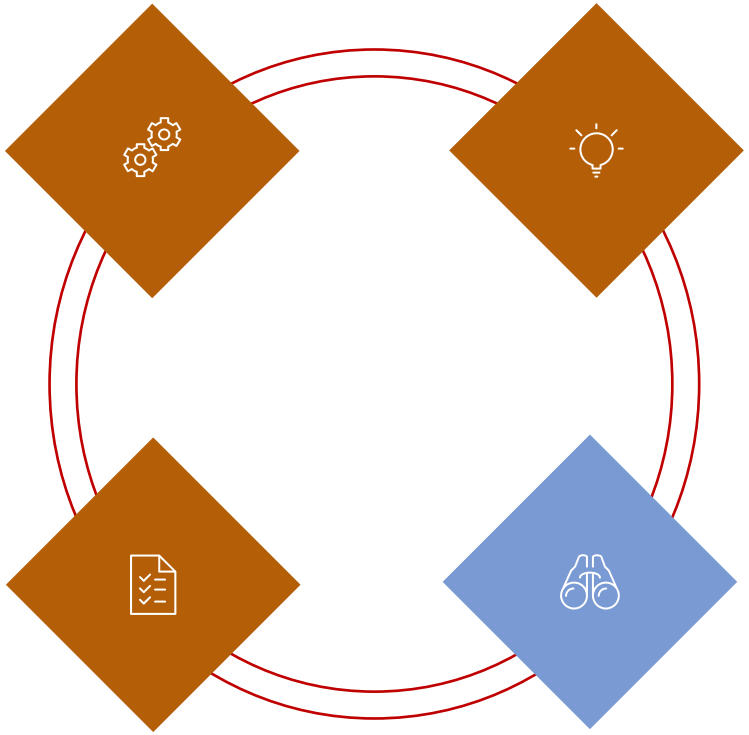
Conclusions & Future Work

Time Reduction

The proposed scheme tries to identify the needs of analytics queries and allocate them to edge nodes that immediately respond and own data that match to their constraints. The aim is to limit the time for the conclusion and the time for getting the final response

Appropriate Selection

We provide performance results related to the optimal selection of the available nodes and show the ability of our scheme to select the appropriate nodes



Throughput Increase

We provide simulation results that reveal the limited time for concluding an allocation leading to the increase of the throughput of the schemes managing the incoming queries

One-one to Many-many

Our future extensions will allow the definition of a scheme that performs allocations in a many-to-many scheme trying to efficiently manage batch of queries arriving in a management entity

Thank You!

Questions?