Scheduling dependent tasks within a smart city's fog/edge infrastructure powered by renewable energy

New Challenges in Scheduling Theory @ Aussois

Miguel Vasconcelos, Postdoc @ IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse,France
Georges Da Costa, IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France
Patricia Stolf, IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France

May 2024

Introduction

- Cloud-edge continuum
  - Edge nodes with lower latency
  - Possible to offload tasks to the cloud
- Renewable energy
  - Low power consumption of edge nodes
- Smart-cities
  - Applications often require low response time

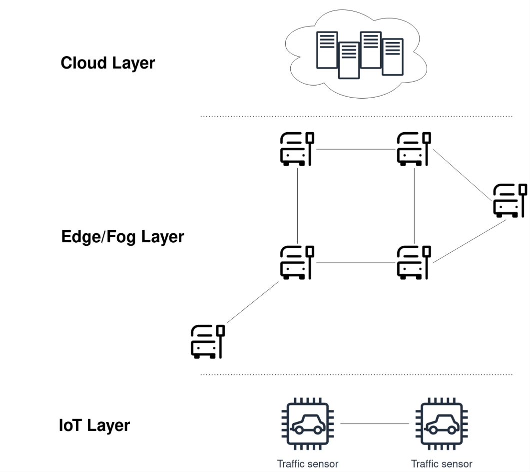

Figure 1: Cloud-Edge continuum in our context.

Introduction

- **VILAGIL** project : improve mobility in Toulouse with smart city approach
- **Opportunistic computing**: Use the computational capacity already present in the city (computers at bus stops, metro stations ...)
- Hosts supplied by renewable energy

Figure 2: Metro computer.

Figure 3: Tramway computer.

Example of user request

Figure 4: Vieille-Toulouse map.

How long it takes from moving between the city ?

- Fuel consumed
- CO_2 emissions
- Costs
- Different modes of transportation

Example of user request

Figure 4: Vieille-Toulouse map.

How long it takes from moving between the city ?

- Fuel consumed
- CO_2 emissions
- Costs
- Different modes of transportation

Traffic prediction : Essential information

Modeling a city

Figure 5: Street graph of Vieille-Toulouse.

The city is represented as a graph of streets

- Edges are the streets
- Vertices are the interconnections between the streets

Example of a route

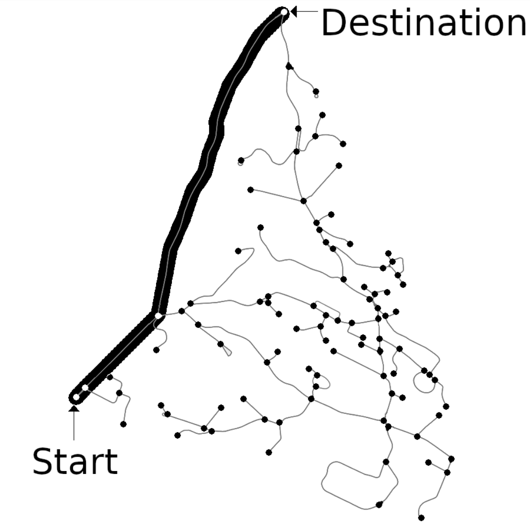

Figure 6: Example of a request.

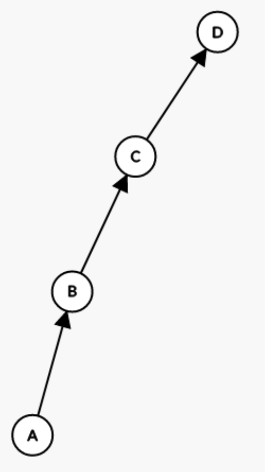

Figure 7: Tasks for the traffic computation.

Example of a route - Impact of the other streets

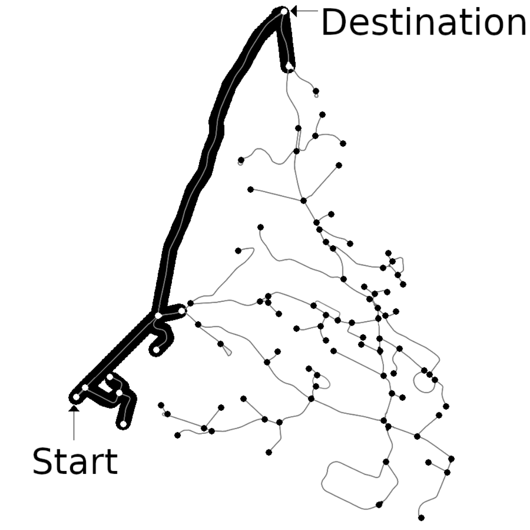

Figure 8: Example of one request.

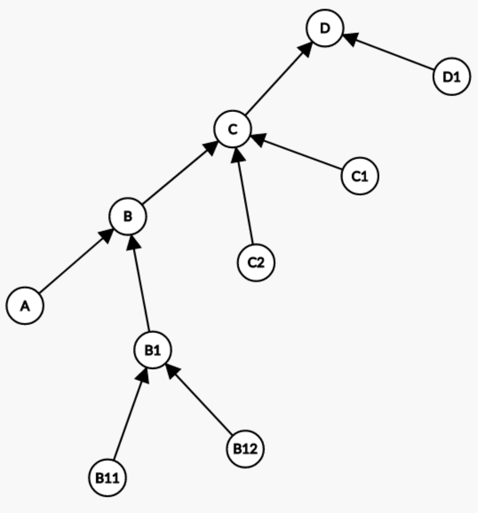

Figure 9: Tasks for the traffic computation.

Example of a route - Source of the data

Figure 10: Example of one request.

- Traffic data is collected from the sensors around the city
- Each bus stop manages the data of the closest streets
- Communication is needed between the tasks to share the data

Example of a route - Source of the data

Figure 11: Example of one request.

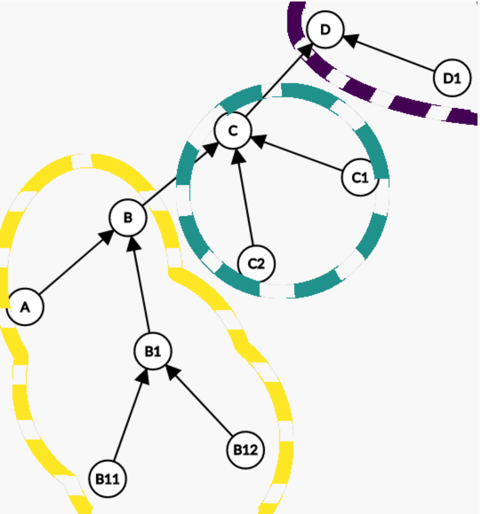

Figure 12: Tasks for the traffic computation.

Requests and task model - Summary

- A user request about a path is represented as a Directed Acyclic Graph (DAG) of tasks
- Each task represents computation for a segment/street (prediction of traffic)
- Each edge/fog node (bus stop) has local information of the streets (nearest streets)

Requests and task model - Summary

- A user request about a path is represented as a Directed Acyclic Graph (DAG) of tasks
- Each task represents computation for a segment/street (prediction of traffic)
- Each edge/fog node (bus stop) has local information of the streets (nearest streets)

How to schedule the tasks to the fog/edge infrastructure aiming to reduce response time and non-renewable energy consumption ?

- Unrelated machines (different computing speeds, power consumption ...)
- Dependent tasks (prec), machine-dependent communication speeds (edge-edge, edge-cloud)

Experiments

Initial Experiments

- Why the opportunistic approach?
  - Comparison between a centralized approach in terms of response time and energy consumption
- How to schedule the workload in the distributed approach?
  - Comparison between different algorithms in terms of response time, energy consumption and renewable energy usage

Experiment design and assumptions

- Computational simulations using the SimGrid framework[1]
- Tasks modeling:
    - One tasks uses 100% of one CPU core
    - 0.1 s if executed in edge node; 0.05 s in the cloud
- Network modeling:
    - Flow-level TCP modeling
    - No bandwidth limitations (TCP slow start not considered)
    - Focus in network latency
- Energy consumption:
    - Linear model based on CPU usage
    - Static part (idle) + dynamic part (based on CPU usage)

[1] Casanova, Henri, Giersch, Arnaud, Legrand, Arnaud, Quinson, Martin, Suter, Frédéric. Versatile, Scalable, and Accurate Simulation of Distributed Applications and Platforms. Journal of Parallel and Distributed Computing, 2014.

Experiment I : Centralized vs distributed

Computational infrastructure modeling

Centralized approach:

- Server with 64 CPU cores
- 66 W when idle; 220 W at 100%

Distributed approach:

- 17 Raspberry PI with 4 CPU cores each
- 2.5 W when idle; 7.3 W at 100%
- (total of 42.5 W when idle, and 124.1 W at 100%)
- Network links latency:
  - 10 ms between edge nodes

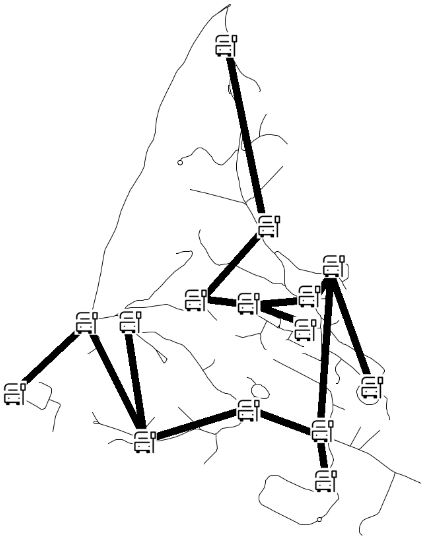

Figure 13: Distributed infrastructure.

Workload

- Inspired by real mobility data [2]
- 30000 requests

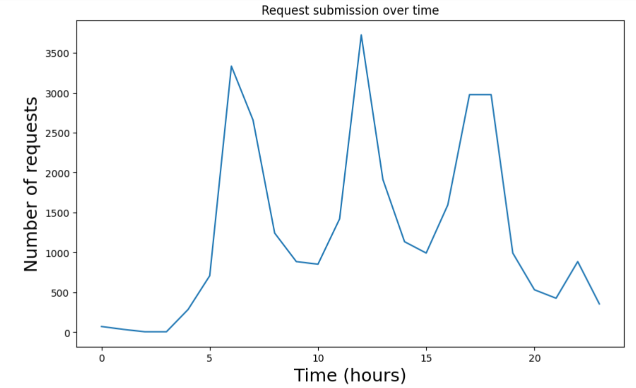

Figure 14: Workload distribution

[2] Metro SP, Pesquisa Origem e Destino 2017.

Results

Using the distributed version:

- **35% of energy savings**
- Average **response time reduced by 40%**
  - Tasks executed closer to where the data is produced

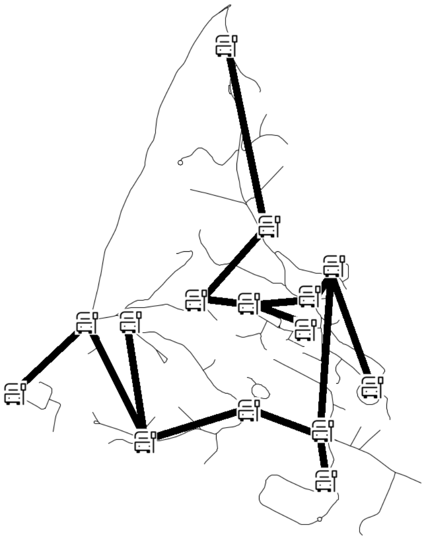

Figure 15: Distributed infrastructure.

Experiment II :Different scheduling algorithms for the distributed approach

Computational infrastructure modeling

- 17 Raspberry PI with 4 CPU cores (fog/edge): 2.5 W when idle; 7.3 W at 100%
- Server with 256 CPU cores (cloud): 66 W when idle; 220 W at 100%
- Network links latency:
  - 10 ms for edge/fog
  - 100 ms for cloud

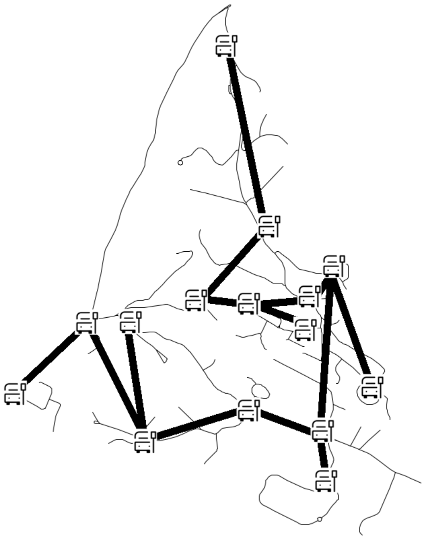

Figure 16: Edge/Fog infrastructure

Algorithms

Baseline

- Allocate to the bus stop that have its required data
- (less communications)

Heterogeneous Earliest Finish Time (HEFT)[3]

- Allocate to the host with that will have the earliest finish time (considering computations and communications)

Green Earliest Finish Time (GEFT)

- Allocate to the host that have green energy and the earliest finish time
- inspired in the HEFT algorithm

[3] Topcuoglu, Haluk; Hariri, Salim; Wu, M. (2002). 'Performance-effective and low-complexity task scheduling for heterogeneous computing". IEEE Transations on Parallel and Distributed Systems. 13 (3): 260–274

Energy modeling

- Edge hosts have PV panels and batteries, and use grid as backup
- Electricity from the grid is assumed to be carbon intensive, and renewable from the cloud layer
- Solar irradiation values per minute (NASA MERRA-2)[3]
- Small variation of irradiation between the different hosts (considering a city)

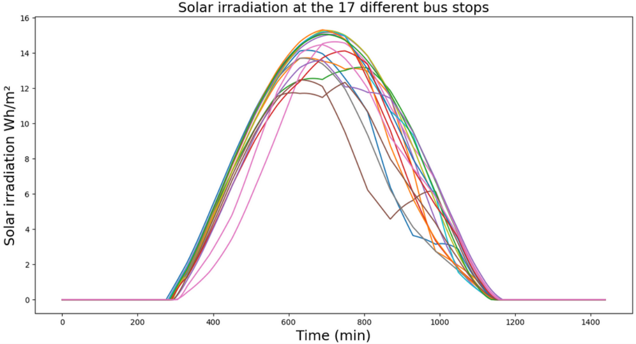

Figure 17: Solar irradiation values

Results - Requests response time

Table 1: Statistics of the request response time (is seconds) by algorithm

Alg.	Mean	Median	90%	95%	99%
Baseline	7.75	1.02	28.15	38.25	56.52
HEFT	0.74	0.73	0.98	0.99	1.07
GEFT	0.82	0.84	1.02	1.04	1.28

Results - Task allocation

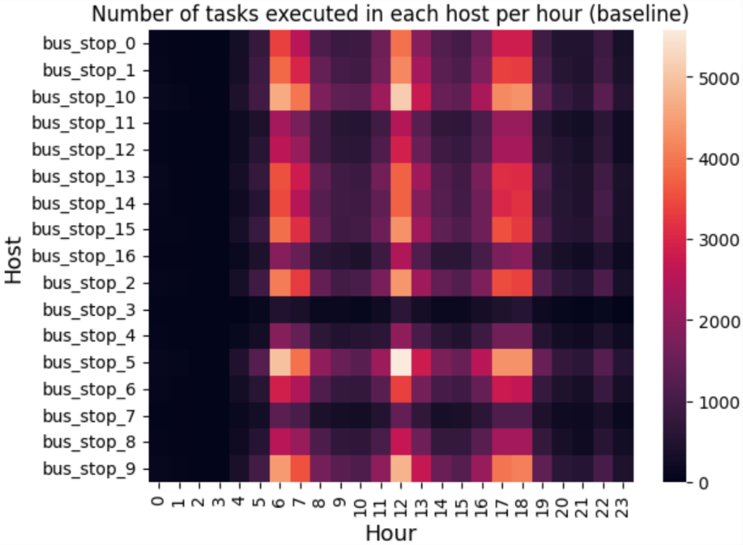

Figure 18: Task allocation over time for the baseline algorithm.

Figure 19: Task allocation over time for the HEFT algorithm.

Results - Task allocation

Figure 20: Task allocation over time for the GEFT algorithm.

Results - Energy consumption

Table 2: Energy consumption by algorithm

Alg.	Total (Wh)	Non-Renewable (Wh)	Renewable energy usage (%)
Baseline	29.87	5.98	80%
HEFT	30.69	3.99	87%
GEFT	36.08	0.08	99.7%

*Results without Idle time

Summary

- Scheduling dependent tasks into a fog/edge infrastructure
- Presence of renewable energy in the hosts
- Improve QoS
- Increase renewable energy usage

Ongoing work

Figure 21: Example of path in Toulouse.

Considering Toulouse

- 1638 hosts (open street map)
    - bus stops
    - tram stops
    - metro stations
    - train stations

Ongoing work

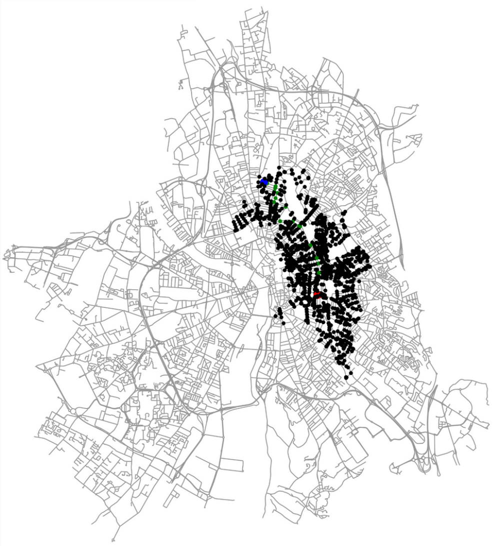

Figure 22: Example of path in Toulouse.

Considering Toulouse

- 1638 hosts (open street map)
  - bus stops
  - tram stops
  - metro stations
  - train stations
- Requests with more tasks to execute: 6949 tasks in the example

Benefits and challenges of simulations

- Execution time of simulations (using a laptop)
    - 2 minutes (centralized, baseline), 8 minutes (GEFT, HEFT) to simulate one day
    - 1 CPU core (possibility to run longer periods of time in parallel)
- Challenges in network modeling :
    - Mobile network modeling, mobility of nodes in space, dynamic latency

Other future Research directions

- Shutdown idle hosts and manage the workload to the other hosts
- Caching
- Other scheduling strategies
- Information of the climate conditions and users requests in the scheduling decision
- Adding new servers in the edge layer and the trade-off between costs ($, CO_2) and QoS

Acknowledgments

- GIS neOCampus of Université Toulouse III Paul Sabatier
- This work was carried out within the MaaS action of the VILAGIL project, a project co-financed by Toulouse Métropole and France 2030 as part of the Territoires d'innovation program operated by the Banque des territoires

Figure 23: Funding projects and agencies

Thank you !

Thank you for your attention!

Contact: miguel-felipe.silva-vasconcelos@irit.fr

Scheduling dependent tasks within a smart city's fog/edge infrastructure powered by renewable energy

Miguel Felipe Silva Vasconcelos

Postdoc @ IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France

Results - Requests response time

Table 3: Statistics of the request response time (is seconds) by algorithm

Alg.	Mean	Median	90%	95%	99%
Baseline	7.75	1.02	28.15	38.25	56.52
Baseline (Vivaldi)	4.20	0.785	14.61	21.48	32.22
Centralized	10.49	1.18	35.16	43.41	60.10
Centralized (Vivaldi)	1.45	0.53	3.055	7.29	13.23

Results - Requests response time

Table 4: Statistics of the request response time (is seconds) by algorithm

Scenario	Mean	Median	90%	95%	99%
Opportunistic	4.20	0.785	14.61	21.48	32.22
Centralized	1.45	0.53	3.055	7.29	13.23