OTEC: An Optimized Transcoding Task Scheduler for Cloud and Fog Environments

Samira Afzal, Farzad Tashtarian, Hamid Hadian, Alireza Erfanian, Christian Timmerer, and Radu Prodan






Institute of Information Technology (ITEC), Alpen-Adria-Universität Austria
samira.afzal@aau.at | https://athena.itec.aau.at/

UNIVERSITÄT KLAGENFURT
FAKULTÄT FÜR TECHNISCHE WISSENSCHAFTEN
INSTITUT FÜR INFORMATIONSTECHNOLOGIE

ATHENA

APOLLO

Agenda

- Motivation and main objectives
- OTEC
- Evaluated results
- Conclusion

Motivation and Main Objectives

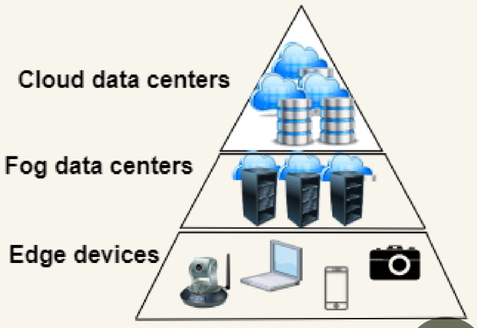
Video streaming is a significant part of the current network traffic

Video encoding is

- Computationally intensive
- Costly
- Time-consuming

To select **Cloud/Fog resources** for video encoding/transcoding operations, aiming at:

- Minimizing cost and/or minimizing encoding time
- Making a trade-off between encoding time and cost
- Considering encoding time deadline and cost limitation



2

Motivation and Main Objectives

Video streaming is a significant part of the current network traffic

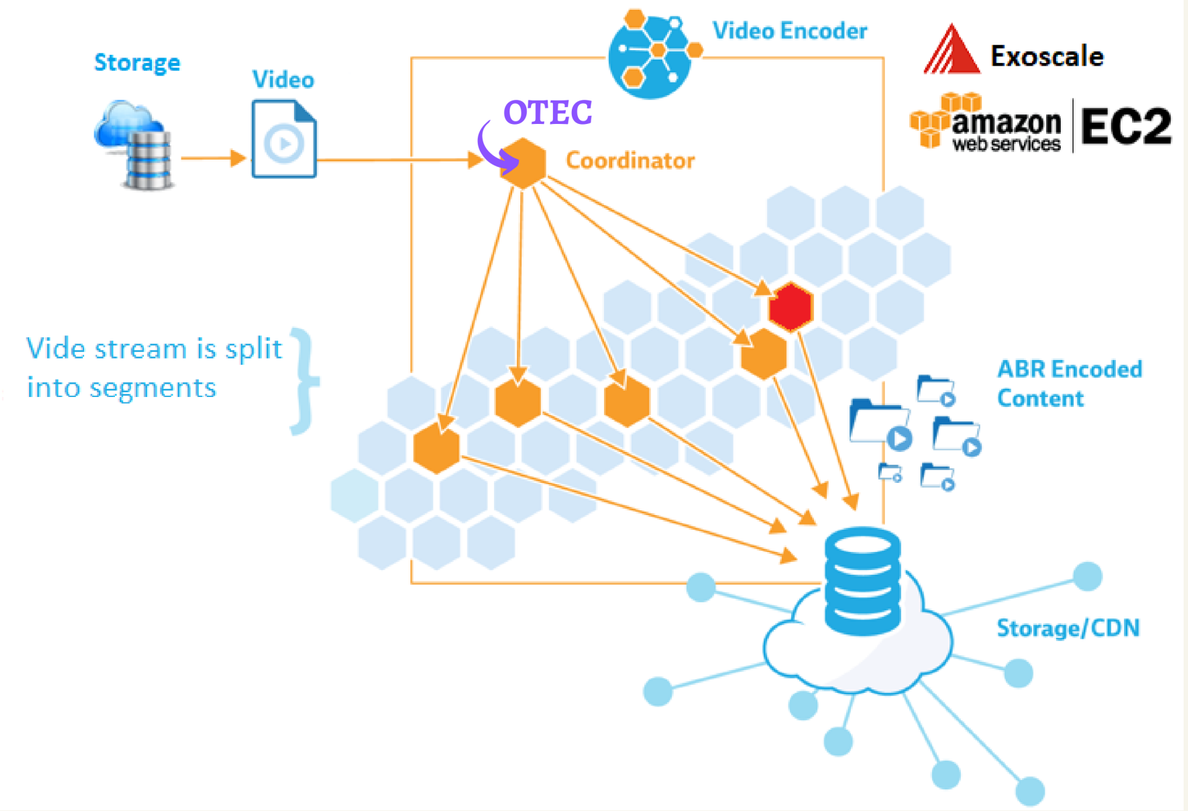
Video encoding is

- Computationally intensive
- Costly
- Time-consuming

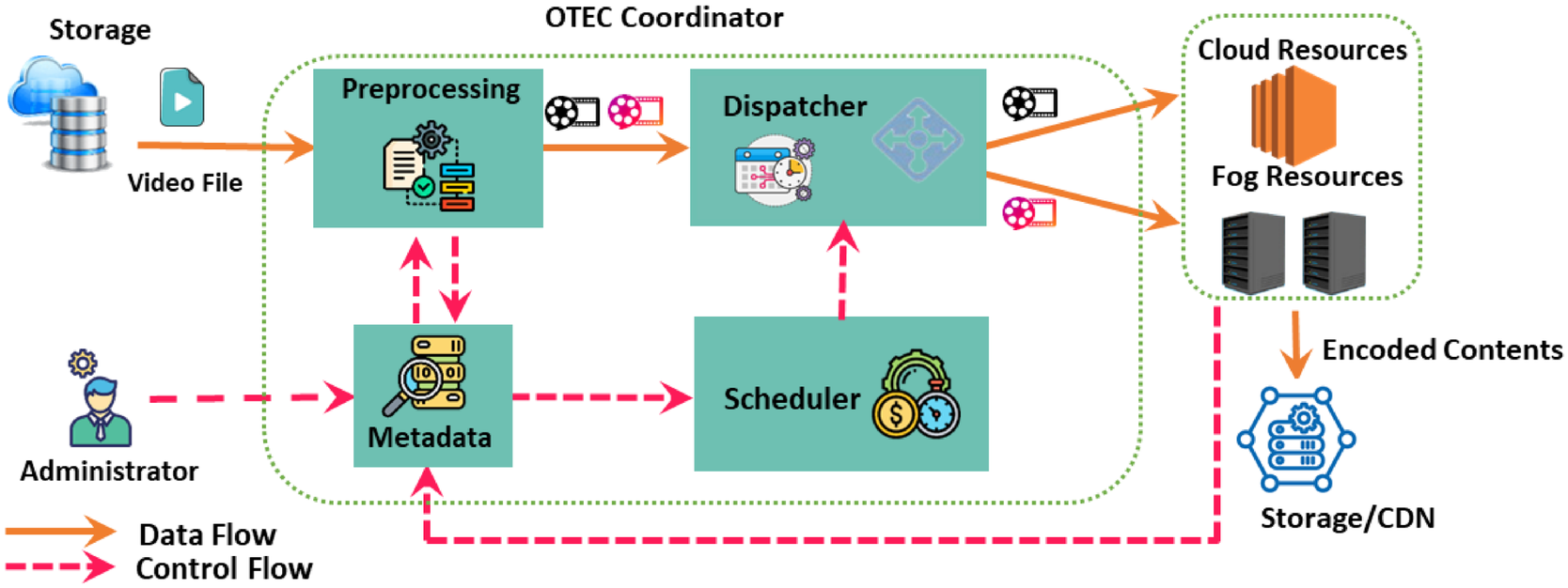
To select **Cloud/Fog resources** for video encoding/transcoding operations, aiming at:

- Minimizing cost and/or minimizing encoding time
- Making a trade-off between encoding time and cost
- Considering encoding time deadline and cost limitation

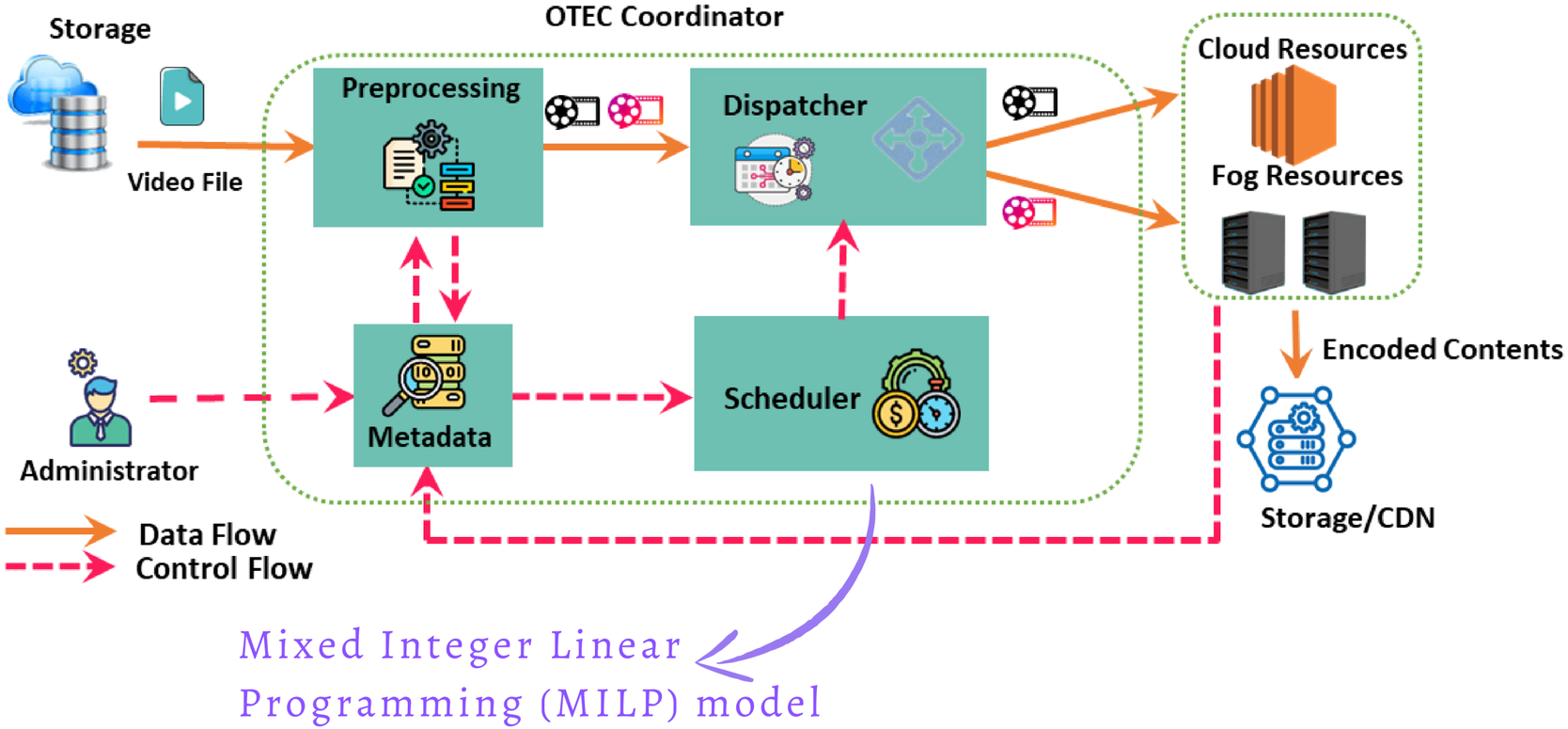
Optimized Encoding Task Scheduler for Cloud and Fog Environments (OTEC)



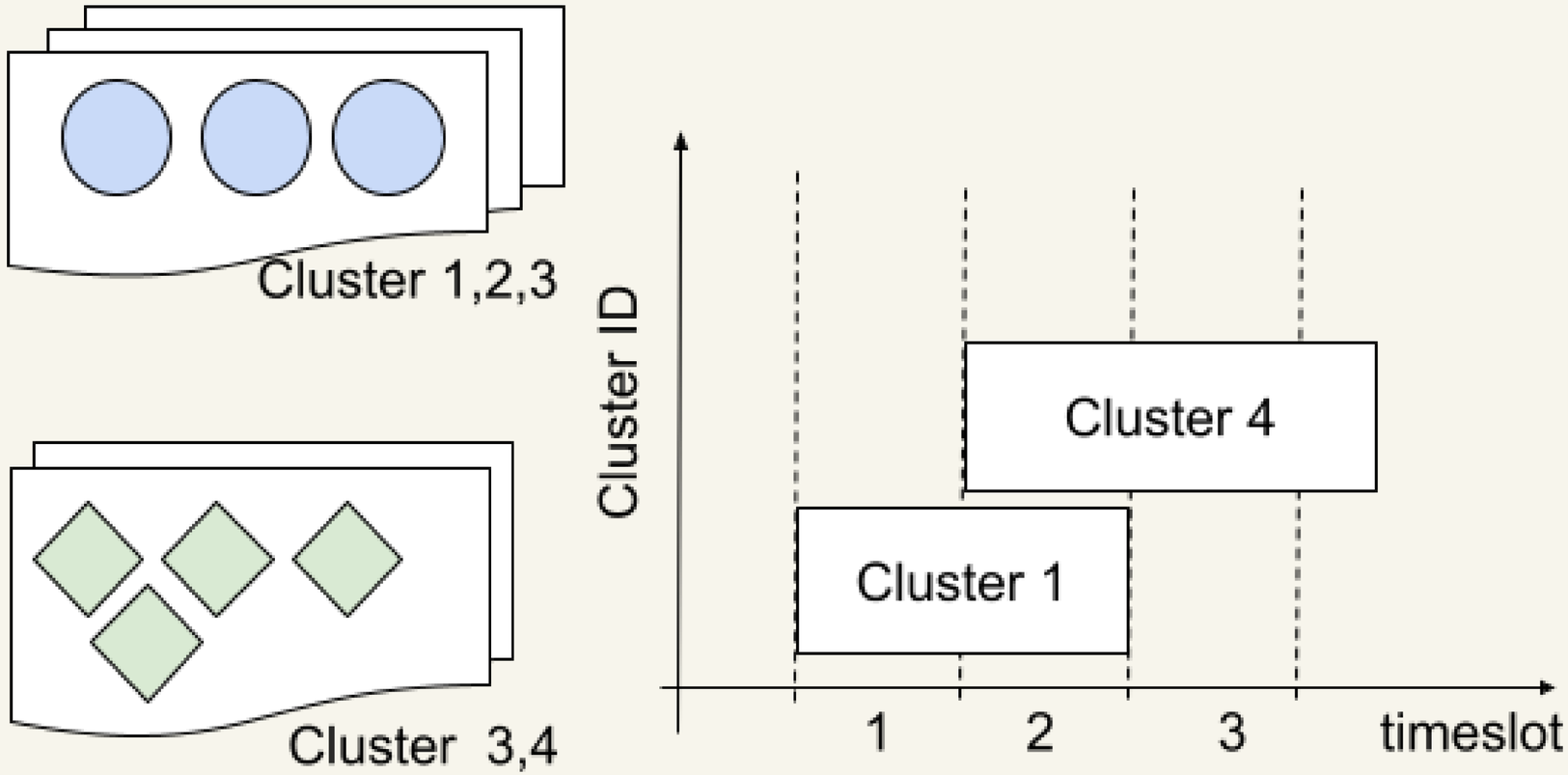
OTEC Architecture Overview



OTEC Architecture Overview



Resource Allocation Example



Optimization Model

Sets of constraints:

- Cluster selection
  - One cluster per timeslot
  - Select the required number of instances
  - Cluster start and end times
  - Non-overlapping encoding timeslots
- Time and cost
  - Total transcoding time
  - Total transcoding cost

Objective Function

To optimize a multi-objective function representing a linear combination of two metrics, **total transcoding cost** and **total transcoding time**

$$\min\left\{\alpha\cdot\frac{\lambda}{\Lambda}+(1-\alpha)\cdot\frac{\theta}{\Theta}\right\}$$

Objective Function

To optimize a multi-objective function representing a linear combination of two metrics, **total transcoding cost** and **total transcoding time**

$$\min\left\{\alpha \cdot \frac{\lambda}{\Lambda} + (1-\alpha) \cdot \frac{\theta}{\Theta}\right\}$$

Maximum possible transcoding time

Maximum possible cost

Administration priority $\in [0, 1]$

Time-optimized (= 0)

Cost-optimized (= 1)

Evaluation

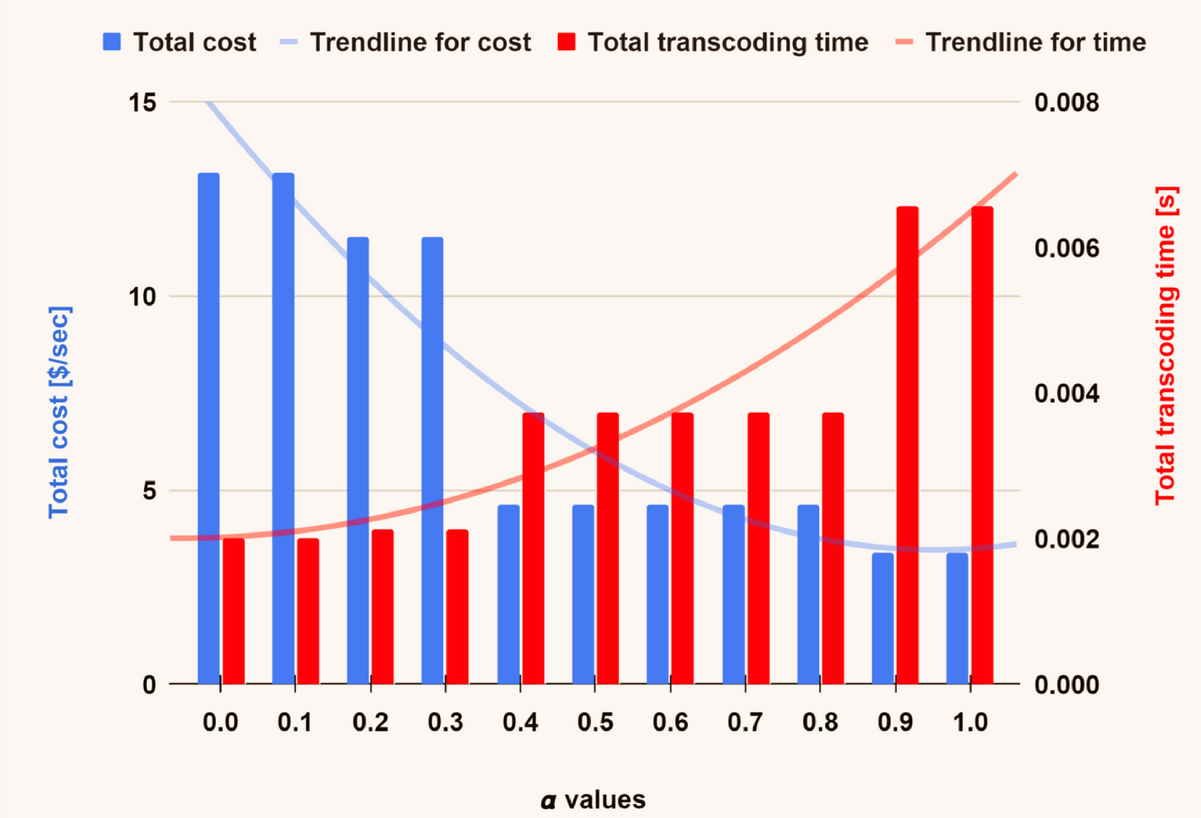
- Various administration priorities
- Various timeslot duration and number of segments
- Various segment transcoding times

Experimental Infrastructure

Provider	Instance type	Instance name	CPU #cores	Memory [GB]	Price [$/hour]	Number of instances
Cloud* (AWS)	Compute optimized	c5.xlarge	4	8	0.195	10
		c5.4xlarge	16	32	0.776	50
		c5.9xlarge	36	72	1.764	60
	General purpose	m5.xlarge	4	16	0.23	10
		m5.2xlarge	8	32	0.46	20
		m5.4xlarge	16	64	0.92	20
	Memory optimized	r5.xlarge	4	32	0.304	30
		r5.2xlarge	8	64	0.608	60
		r5.4xlarge	16	128	1.216	30
Fog** (Exoscale)	Standard	large	4	8	0.088	10
		huge	8	32	0.355	20
	CPU optimized	extra large	16	32	0.444	50

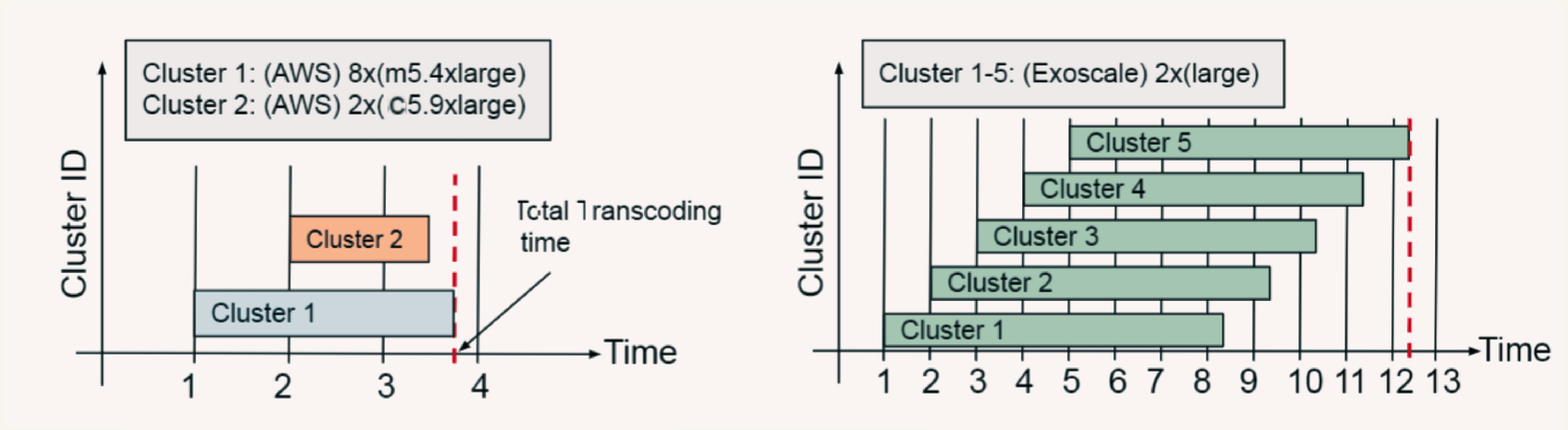
*https://aws.amazon.com/ec2/pricing/on-demand/
**https://www.exoscale.com/pricing/

Administration Priority



- Time-optimized results are **30.5% faster** and **24.78% more expensive** than the cost-optimized ones.
- For = 0.5, there is a tradeoff between the total cost (0.0024 $/s) and transcoding time (6.99 s).

Resource Allocation Results



Time-optimized ($\alpha = 0$) **Cost-optimized ($\alpha = 1$)**

- Coordinator on a local cloud instance
  - A four-core Intel core-i7 processor
  - 32 GB of memory
- 10 segments

Timeslot Duration and Number of Segments

Timeslot [s]	1	2	4	6	8	8	8
Number of segments	10	10	10	10	10	20	200
Number of clusters	117	68	39	22	18	18	18
Number of constraints	8797	5122	2947	1672	1372	1372	1372
Number of variables	3512	2042	1172	662	542	542	542
Scheduling time [s]	9.47	9.915	1.54	1.74	0.82	1.23	0.97
Total transcoding time [s]	3.76	4.76	6.8	8.42	9.42	13.3788	49.42
Total cost [$/sec]	0.007	0.0105	0.0208	0.0052	0.0069	0.0517	0.1658

- Time-optimized scenario
- Coordinator on a local cloud instance
  - A four-core Intel core-i7 processor
  - 32 GB of memory
- 370 instances

Timeslot Duration and Number of Segments

Timeslot [s]	1	2	4	6	8	8	8
Number of segments	10	10	10	10	10	20	200
Number of clusters	117	68	39	22	18	18	18
Number of constraints	8797	5122	2947	1672	1372	1372	1372
Number of variables	3512	2042	1172	662	542	542	542
Scheduling time [s]	9.47	9.915	1.54	1.74	0.82	1.23	0.97
Total transcoding time [s]	3.76	4.76	6.8	8.42	9.42	13.3788	49.42
Total cost [$/sec]	0.007	0.0105	0.0208	0.0052	0.0069	0.0517	0.1658

- Time-optimized scenario
- Coordinator on a local cloud instance
  - A four-core Intel core-i7 processor
  - 32 GB of memory
- 370 instances

Timeslot Duration and Number of Segments

Timeslot [s]	1	2	4	6	8	8	8
Number of segments	10	10	10	10	10	20	200
Number of clusters	117	68	39	22	18	18	18
Number of constraints	8797	5122	2947	1672	1372	1372	1372
Number of variables	3512	2042	1172	662	542	542	542
Scheduling time [s]	9.47	9.915	1.54	1.74	0.82	1.23	0.97
Total transcoding time [s]	3.76	4.76	6.8	8.42	9.42	13.3788	49.42
Total cost [$/sec]	0.007	0.0105	0.0208	0.0052	0.0069	0.0517	0.1658

- Time-optimized scenario
- Coordinator on a local cloud instance
  - A four-core Intel core-i7 processor
  - 32 GB of memory
- 370 instances

Timeslot Duration and Number of Segments

Timeslot [s]	1	2	4	6	8	8	8
Number of segments	10	10	10	10	10	20	200
Number of clusters	117	68	39	22	18	18	18
Number of constraints	8797	5122	2947	16[illegible]2	1372	1372	1372
Number of variables	3512	2042	1172	66[illegible]	542	542	542
Scheduling time [s]	9.47	9.915	1.54	1.74[illegible]	0.82	1.23	0.97
Total transcoding time [s]	3.76	4.76	6.8	8.42	9.42	13.3788	49.42
Total cost [$/sec]	0.007	0.0105	0.0208	0.0052	0.0069	0.0517	0.1658

15% ⬇

~2.5 times ⬆

- Time-optimized scenario
- Coordinator on a local cloud instance
  - A four-core Intel core-i7 processor
  - 32 GB of memory
- 370 instances

Timeslot Duration and Number of Segments

Timeslot [s]	1	2	4	6	8	8	8
Number of segments	10	10	10	10	10	20	200
Number of clusters	117	68	39	22	18	18	18
Number of constraints	8797	5122	2947	1672	1372	1372	1372
Number of variables	3512	2042	1172	662	542	542	542
Scheduling time [s]	9.47	9.915	1.54	1.74	0.82	1.23	0.97
Total transcoding time [s]	3.76	4.76	6.8	8.42	9.42	13.3788	49.42
Total cost [$/sec]	0.007	0.0105	0.0208	0.0052	0.0069	0.0517	0.1658

- Time-optimized scenario
- Coordinator on a local cloud instance
  - A four-core Intel core-i7 processor
  - 32 GB of memory
- 370 instances

Segment Transcoding Time

Coordinator

Segment transcoding time [s]	4.76	10	20	30	40
Total transcoding time [s]	3.760	6.8	11	14	17
Total cost [$/sec]	0.007	0.0128	0.0216	0.038	0.0582
No. of clusters x No. of instances x (Instance type)	1 x 8 x (m5.4xlarge) 1 x 2 x (c5.9xlarge)	1 x 8 x (m5.4xlarge), 1 x 2 x (extra large)	3 x 2 x (c5.9xlarge) 1 x 2 x (c5.4xlarge) 1 x 2 x (extra large)	1 x 2 x (extra large) 4 x 2 x (c5.9xlarge)	5 x 2 x (c5.9xlarge)

4.5 times ↑
8.31 times ↑

- Time-optimized scenario
- Coordinator on a local cloud instance
  - A four-core Intel core-i7 processor
  - 32 GB of memory
- 370 instances
- 10 segments

Conclusions

- **Proposed OTEC for optimized scheduling of adaptive transcoding processes in Cloud and Fog environments**
- Proposed an **MILP** optimization model
- Highlighted factors:
  - **Administrator priorities**, such as cost and/or time priorities
  - **Resources properties**, location, type, cost, number of instances for each resource type
  - **Segment properties**, such as duration, number of segments, segment transcoding time
- Evaluated OTEC on a set of **Cloud AWS** and **Fog Exoscale** resource instances
- Achieved that time-optimized scenarios are **30.5% faster** and **24.78% more expensive** than the cost-optimized ones

Thank you

Have a great day ahead!

Institute of Information Technology (ITEC) Alpen-Adria-Universität Austria
samira.afzal@aau.at https://itec.aau.at/

UNIVERSITÄT KLAGENFURT
FAKULTÄT FÜR TECHNISCHE WISSENSCHAFTEN
INSTITUT FÜR INFORMATIONSTECHNOLOGIE

ATHENA

APOLLO