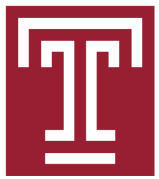
HeteroScore: Evaluating and Mitigating Cloud Security Threats Brought by Heterogeneity

Chongzhou Fang, Najmeh Nazari, Behnam Omidi, Han Wang, Aditya Puri, Manish Arora, Setareh Rafatirad, Houman Homayoun, and Khaled N. Khasawneh

March 2, 2023



LearnDesk

Contents

Introduction

Motivation

Clouds are becoming increasingly heterogeneous

1. New applications being invented
2. New devices being introduced
3. Performance-cost trade-off
4. ...

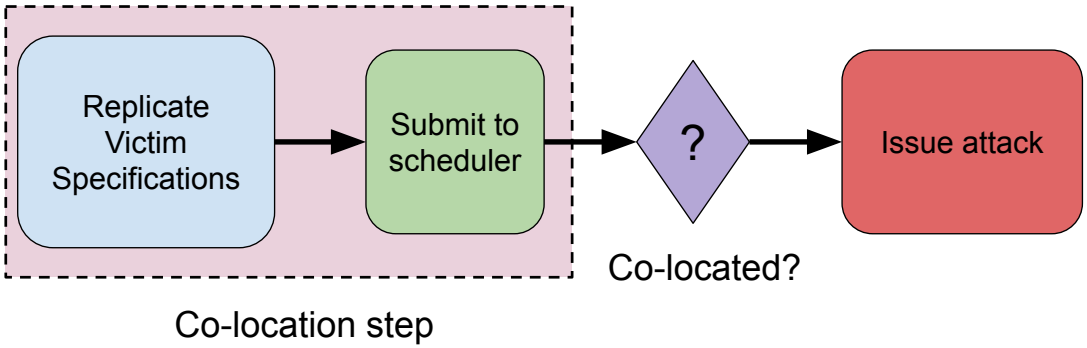
Motivation

Micro-architectural attacks have become a threat to cloud users

1. Side-channel attack
2. Transient execution attack
3. Rowhammer attack
4. ...

A Recap of REPTTACK[1] (NDSS'22)

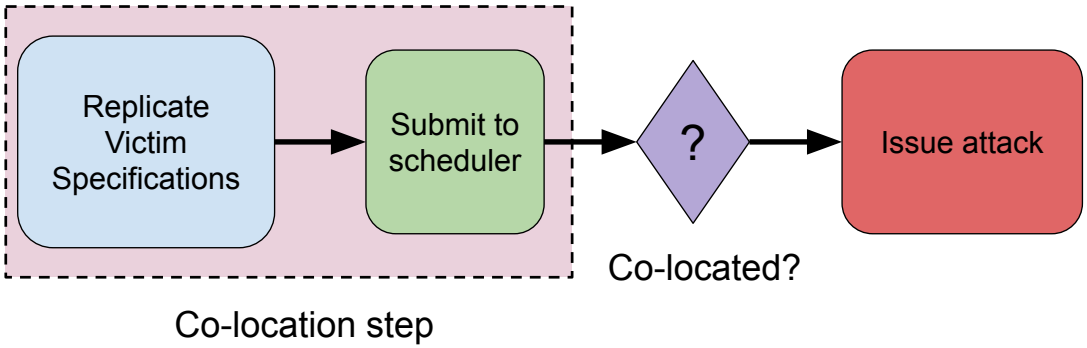
Co-location: an important prerequisite of micro-architectural attacks



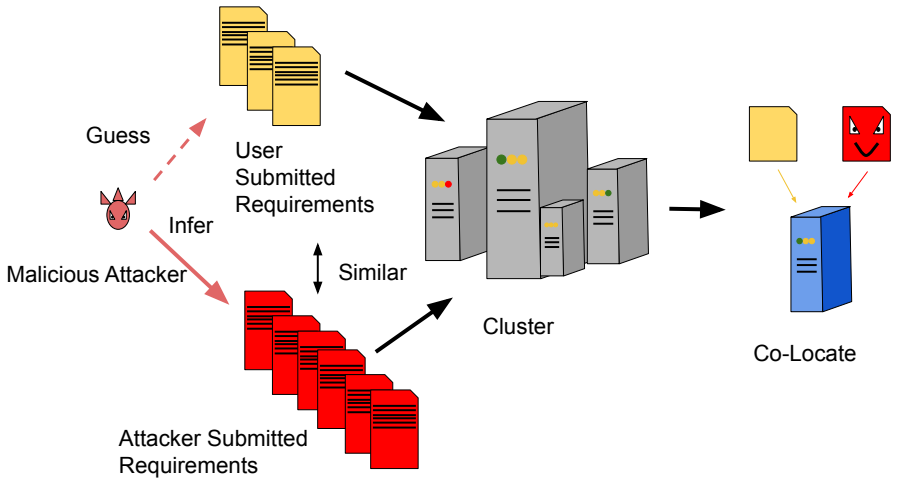
[1] Fang, Chongzhou, et al. "REPTTACK: Exploiting Cloud Schedulers to Guide Co-Location Attacks." NDSS'22.

A Recap of REPTTACK[1] (NDSS'22)

Co-location: an important prerequisite of micro-architectural attacks



REPTTACK[1]



[1] Fang, Chongzhou, et al. "REPTTACK: Exploiting Cloud Schedulers to Guide Co-Location Attacks." NDSS'22.

Motivation: to Quantitatively Measure Security Threats

How insecure is your cluster when facing this kind of attack?

We need a quantitative metric that can:

1. reflect the heterogeneity of a cluster
2. guide defence strategy design

Motivation: to Quantitatively Measure Security Threats

How insecure is your cluster when facing this kind of attack?

We need a quantitative metric that can:

1. reflect the heterogeneity of a cluster
2. guide defence strategy design

Contributions:

1. Heterogeneity Score (HeteroScore)
2. Scheduler-level mitigation technologies inspired by HeteroScore

Threat Model

Threat Model

Cloud Provider

- Neutral

Threat Model

Cloud Provider

- Neutral

Attacker

- Can only perform actions like non-malicious users
- Goal: co-locate with a specific target victim instance

Threat Model

Cloud Provider

- Neutral

Attacker

- Can only perform actions like non-malicious users
- Goal: co-locate with a specific target victim instance

Our Focus

Only on scheduler level, not on hardware level

HeteroScore

Definition and Explanation of HeteroScore

Node representation: multiple 'label-value' description (d-dimension here)

$$\mathrm{N}^{(i)} = (x_1^{(i)}, x_2^{(i)}, ..., x_k^{(i)}, ..., x_d^{(i)})^{\mathrm{T}}$$

Cluster representation (n servers in the cluster):

$$\mathcal{C} = \{\mathrm{N}^{(1)}, \mathrm{N}^{(2)}, ..., \mathrm{N}^{(i)}, ..., \mathrm{N}^{(n)}\}$$

Definition and Explanation of HeteroScore

HeteroScore calculation:

$$\mathcal{H}_c = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{I}\{\rho(\mathrm{N}^{(i)},\mathrm{N}^{(j)}) \leq t_h\}}{n^2}$$

where

$$\rho(\mathrm{N}^{(i)},\mathrm{N}^{(j)}) = \sqrt{\sum_{k=1}^{d}(x_k^{(i)} - x_k^{(j)})^2}$$

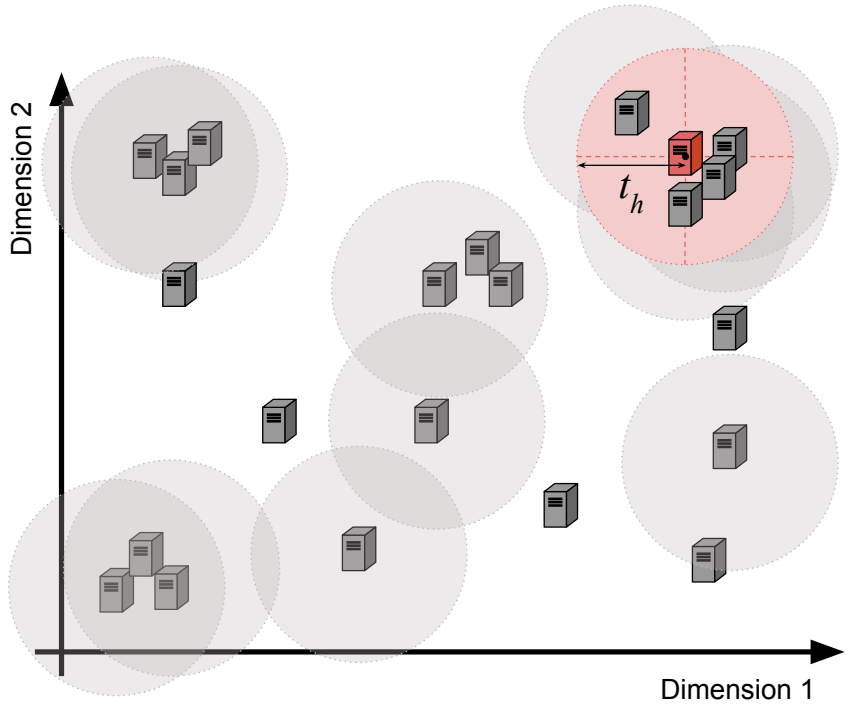
and

$$\mathbf{I}\{*\} = \begin{cases} 1, & \text{Given condition } * \text{ is satisfied,} \\ 0, & \text{Otherwise.} \end{cases}$$

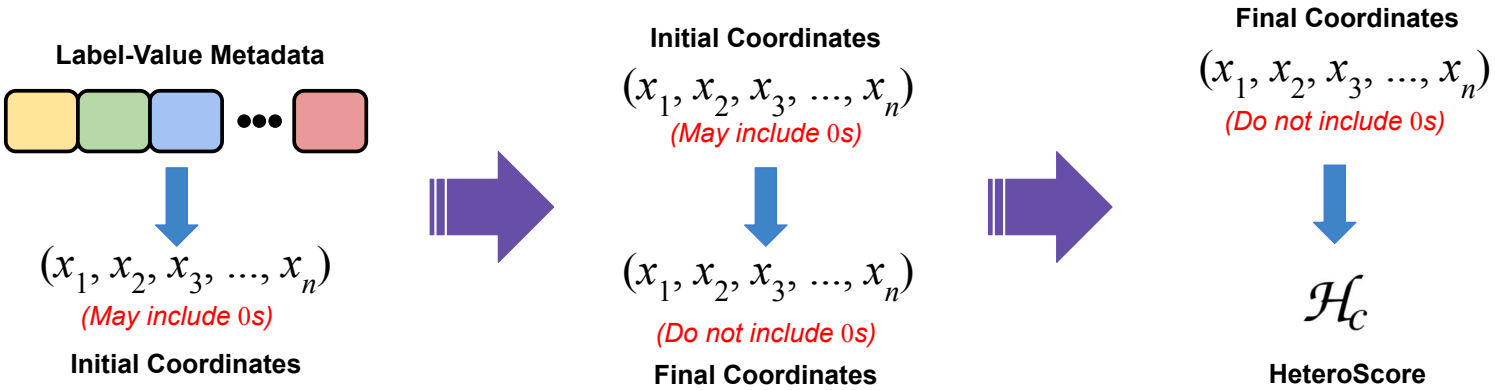
Definition and Explanation of HeteroScore

$\mathcal{H}_c$: Depicts the sparsity of $\mathcal{C}$

$$\mathcal{H}_c = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{I}\{\rho(\mathrm{N}^{(i)}, \mathrm{N}^{(j)}) \leq t_h\}}{n^2}$$



Algorithms


Label-Value Metadata
$(x_1, x_2, x_3, \ldots, x_n)$
(May include 0s)
Initial Coordinates
Initial Coordinates
$(x_1, x_2, x_3, \ldots, x_n)$
(May include 0s)
$(x_1, x_2, x_3, \ldots, x_n)$
(Do not include 0s)
Final Coordinates
Final Coordinates
$(x_1, x_2, x_3, \ldots, x_n)$
(Do not include 0s)
$\mathcal{H}_C$
HeteroScore

Scheduler-Level Mitigation Inspired by HeteroScore

Hiding Label Defence (HLD)

Hiding certain labels from users during scheduling.

Scheduler-Level Mitigation Inspired by HeteroScore

Hiding Label Defence (HLD)

Hiding certain labels from users during scheduling.

Randomly Hiding Label Defence (R-HLD)

Randomly selecting labels to hide from users during scheduling.

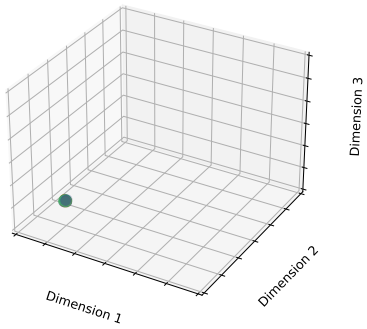
Evaluation

Visualization Results

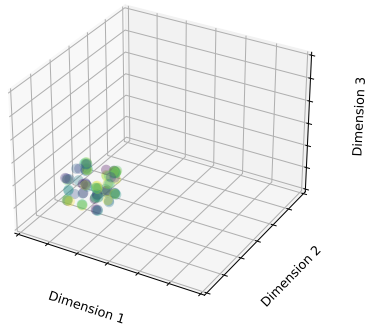
Simulation Environment Setting

- Randomly generated cluster
- n_l: #. of label-value pairs
- n_c: #. of potential choices in each pair

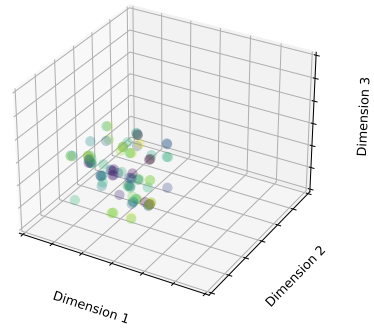
Visualization Results



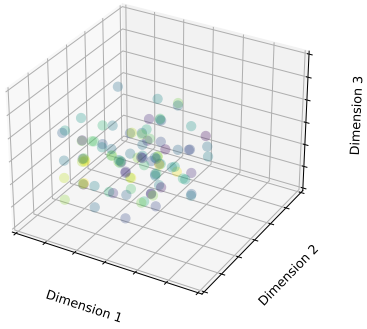
(a) $n_c = 1$, $\mathcal{H}_c = 0$.



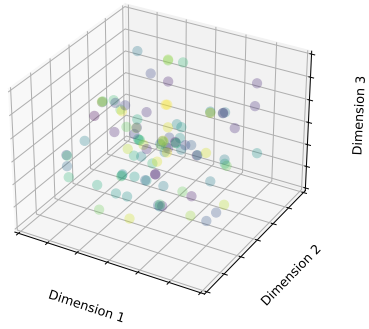
(b) $n_c = 2$, $\mathcal{H}_c = 0.6976$.



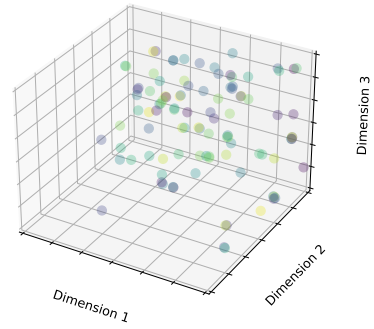
(c) $n_c = 3$, $\mathcal{H}_c = 0.8826$.



(d) $n_c = 4$, $\mathcal{H}_c = 0.9456$.

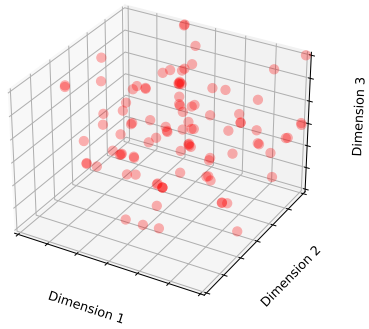


(e) $n_c = 5$, $\mathcal{H}_c = 0.9678$.

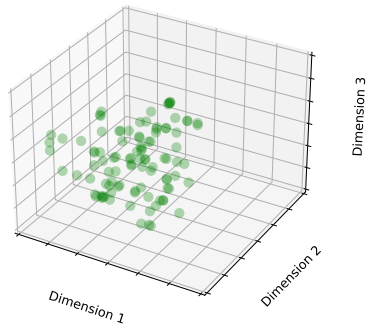


(f) $n_c = 6$, $\mathcal{H}_c = 0.9704$.

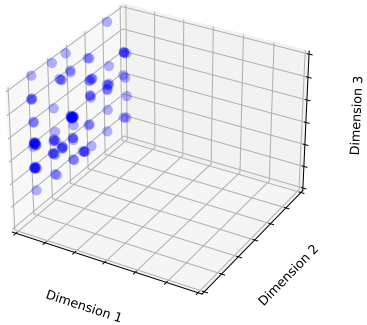
Visualization of Defence



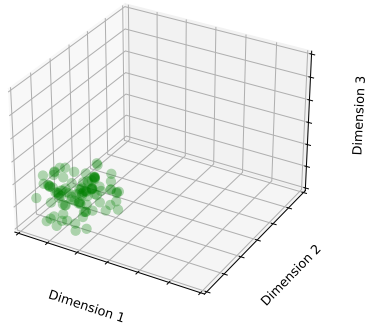
(a) Original, 0.9768.



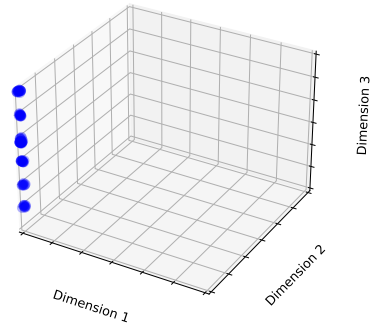
(b) R-HLD, 0.9108.



(c) HLD, 0.915.



(d) R-HLD, 0.6784.



(e) HLD, 0.7118.

HeteroScore Results in Clusters

Simulated Cluster Settings

- A Python simulator mimicking the scheduling policies of Kubernetes
- Nodes are randomly generated

HeteroScore Results in Clusters

Simulated Cluster Settings

- A Python simulator mimicking the scheduling policies of Kubernetes
- Nodes are randomly generated

Physical Cluster Settings

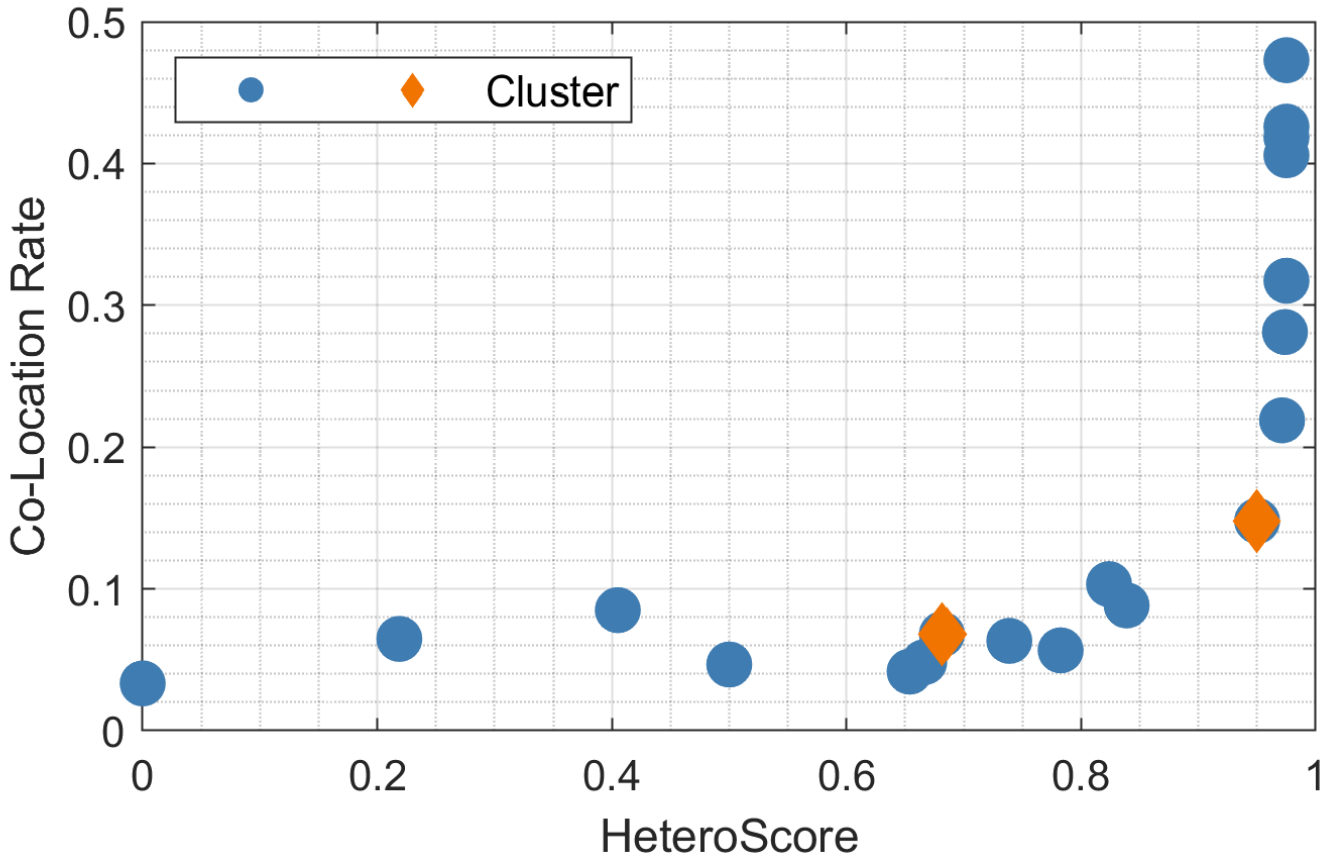
- 40-node Kubernetes cluster in CloudLab

Simulator Results

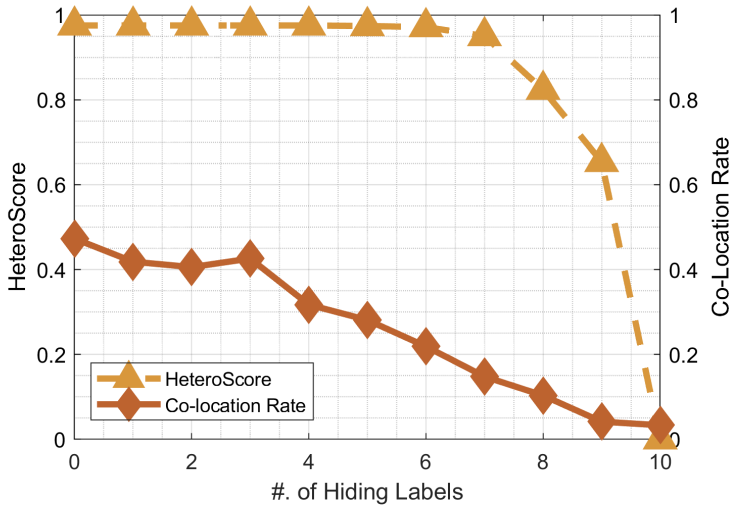
Table: Co-location rates for varying cluster sizes and degree of heterogeneity.

#. of Nodes	$\mathcal{H}_c$	Co-location Rate	
		1-Instance Attack	10-Instance Attack
100	0.9878	51.16%	92.65%
	0.9497	34.04%	65.88%
	0.7126	11.10%	37.42%
	0.4070	4.07%	26.33%
	0	1.12%	8.09%
1,000	0.9975	41.53%	79.20%
	0.9522	15.89%	37.30%
	0.7381	13.78%	22.74%
	0.4084	7.74%	12.35%
	0	1.90%	3.23%
10,000	0.9988	19.88%	65.23%
	0.9437	14.06%	44.09%
	0.7335	7.33%	28.81%
	0.4138	6.42%	9.40%
	0	0.80%	0.87%

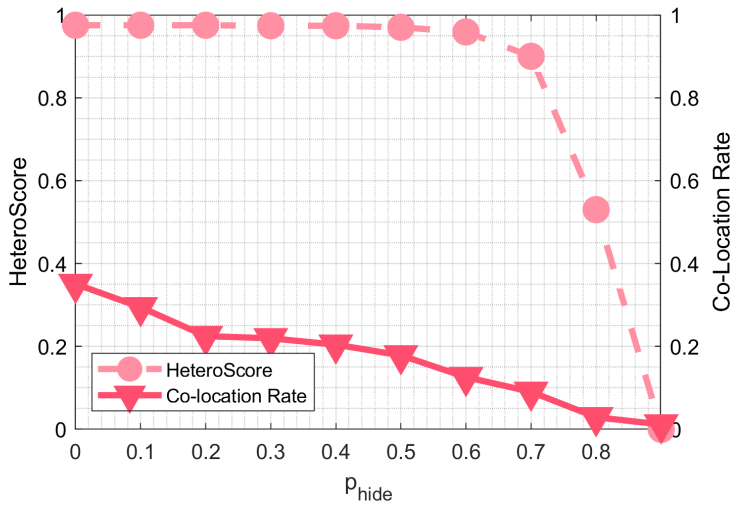
Results in Physical Clusters



Results of HLD & R-HLD



(a) Results of applying HLD.



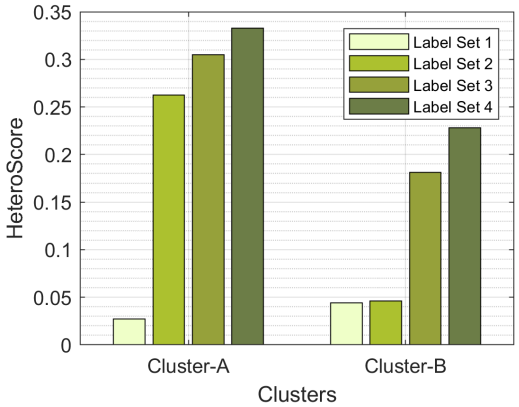
(b) Results of applying R-HLD.

Case Study of University Clusters

Cluster Settings

- University-scale computing clusters managed by SLURM
- Cluster A: 73 servers Cluster B: 194 servers

HeteroScore Calculation



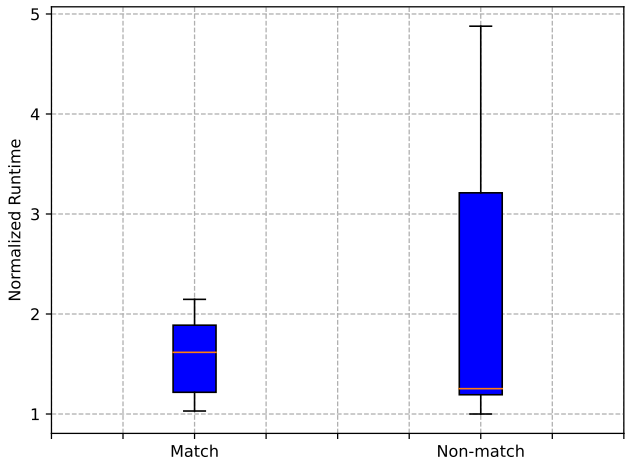
Cluster	Label Set	Labels
Cluster A	1	Partition: Low, Partition: Med
	2	GPU-related labels
	3	Partition: High
	4	Partition: Low, Partition: Med, Partition: High, GPU-related labels
Cluster B	1	Partitions
	2	Partitions, GPU
	3	Bandwidth
	4	Partitions, Bandwidth, GPU

Cost Analysis

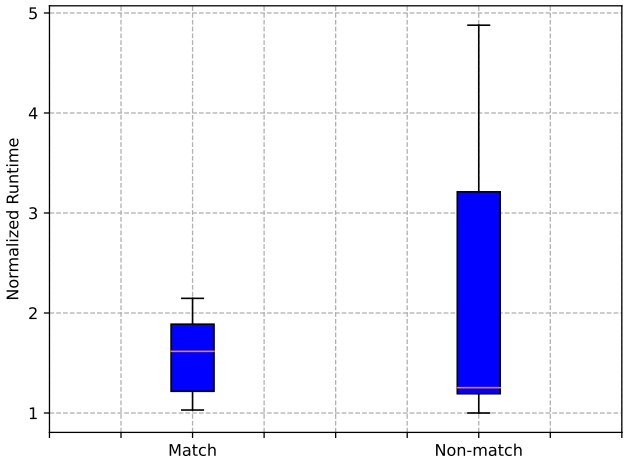
Benchmarks

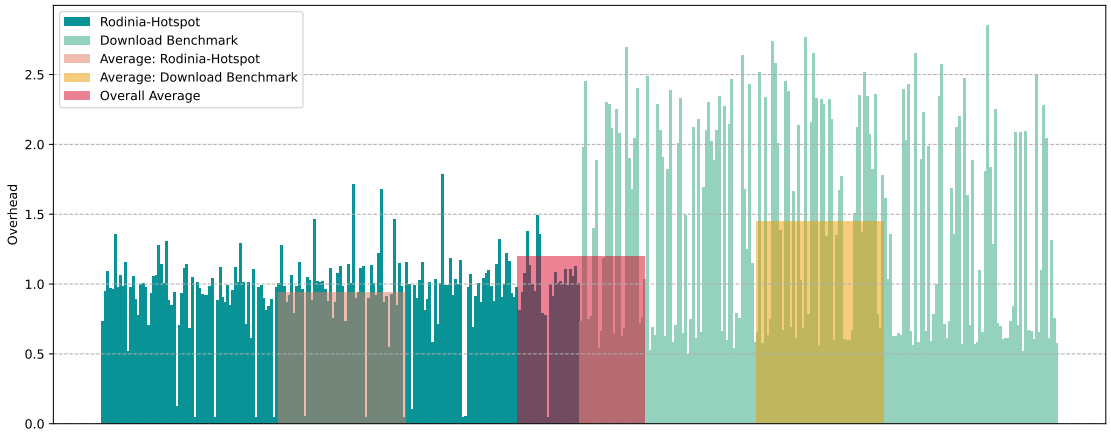
- Network benchmark: downloads contents of specific sizes from the Internet
- Rodinia-Hotspot

Cost Analysis

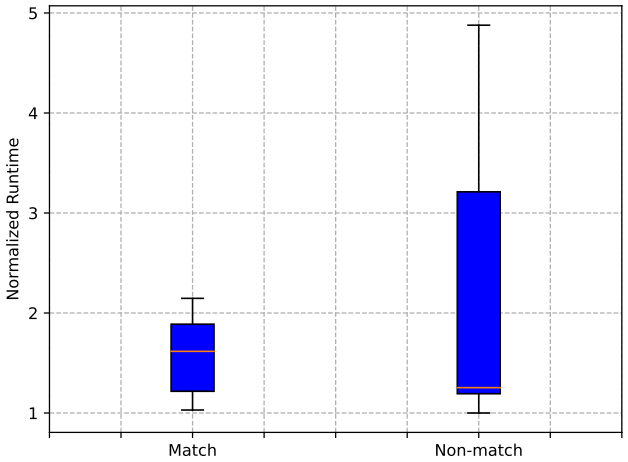


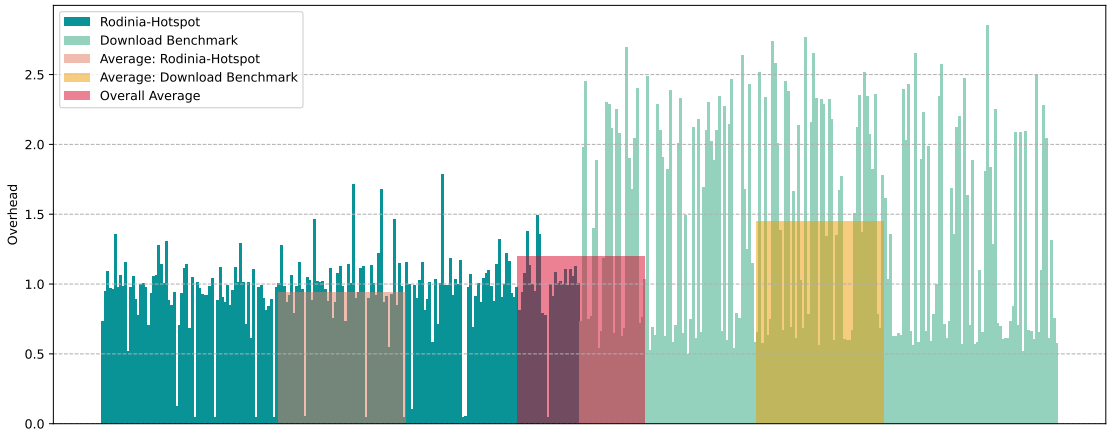
Cost Analysis

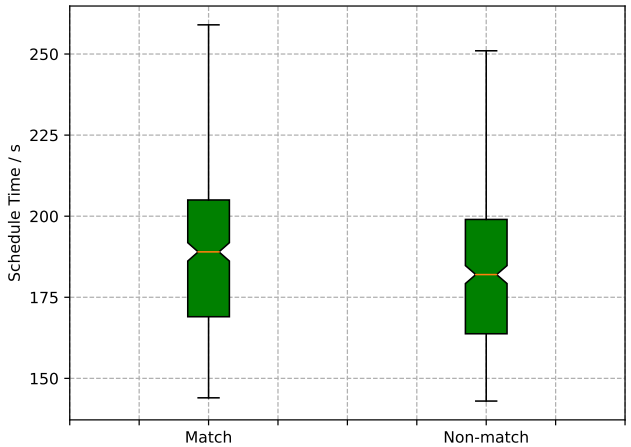




Cost Analysis







Discussion

Choices of Mitigation Strategies

HLD

- More controllable
- Cost more deterministic

Choices of Mitigation Strategies

HLD

- More controllable
- Cost more deterministic

R-HLD

- Cost more balanced

Choices of Mitigation Strategies

HLD

- More controllable
- Cost more deterministic

R-HLD

- Cost more balanced

Combining both strategies

- Selecting a subset of labels to apply R-HLD
- Applying R-HLD with non-uniform parameters

Conclusion

Conclusion

A Metric

- Quantitatively measures the heterogeneity of a cluster
- Can be linked to co-location security

Conclusion

A Metric

- Quantitatively measures the heterogeneity of a cluster
- Can be linked to co-location security

Mitigation Technologies

- HLD
- R-HLD

Thank you!