On Designing Adaptive Data Structures with Adaptive Data “Sub”-Structures

MCS Thesis Defense

Ekaba Bisong

Supervisor: B. John Oommen, Chancellor’s Professor.

Thursday 11th October, 2018

School of Computer Science

Research Outline

- Adaptive Data Structures (ADS) for learning query accesses in Non-stationary Environments (NSEs).
- Survey of available approaches.
- Hierarchical Singly-linked lists (SLLs) on SLLs.
- Object Migration Automaton (OMA) to capture “Locality of Reference”.
- Extended OMA machines.
- Further Research.

Thesis Overview

1. Problem statement and objectives (**Chapter 1**).
2. Survey of Learning Automata (LA), ADS, etc (**Chapter 2**).
3. ADS: Hierarchical SLLs (**Chapter 3**).
4. Enhanced-OMA (EOMA) Hierarchical SLLs (**Chapter 4**).
5. Advanced-OMA (PEOMA, TPEOMA) Hierarchical SLLs (**Chapter 5**).
6. Summary and Conclusion (**Chapter 6**).
7. Additional Results OMA-based schemes (**Appendices**).

Achievements

The results of this research are three potential conference papers:

- Bisong, E. and Oommen, B.J., On Utilizing Enhanced Object Partitioning for Optimizing Self-Organizing Lists in Environments with Locality of Reference. (To be Submitted)
- Bisong, E. and Oommen, B.J., On Utilizing Pursuit-Enhanced Object Partitioning for Optimizing Self-Organizing Lists in Environments with Locality of Reference. (To be Submitted)
- Bisong, E. and Oommen, B.J., On Utilizing Transitivity and Pursuit-Enhanced Object Partitioning for Optimizing Self-Organizing Lists in Environments with Locality of Reference. (To be Submitted)

Overview

Introduction

- The impact of data storage and processing.
- The twin pillars of computing.
- Critical drivers for advances in Artificial Intelligence.
- Goal: Enhance speed of data retrieval.

Non-stationary Environments (NSEs)

- NSEs: Settings that change with time.
- Learning schemes with fixed policies may become non-expedient over time.
- Models of NSEs:
  - Markovian Switching Environments (MSEs).
  - Periodic Switching Environments (PSEs).
- Query Generators: Simulating dependence models.

Learning Automata

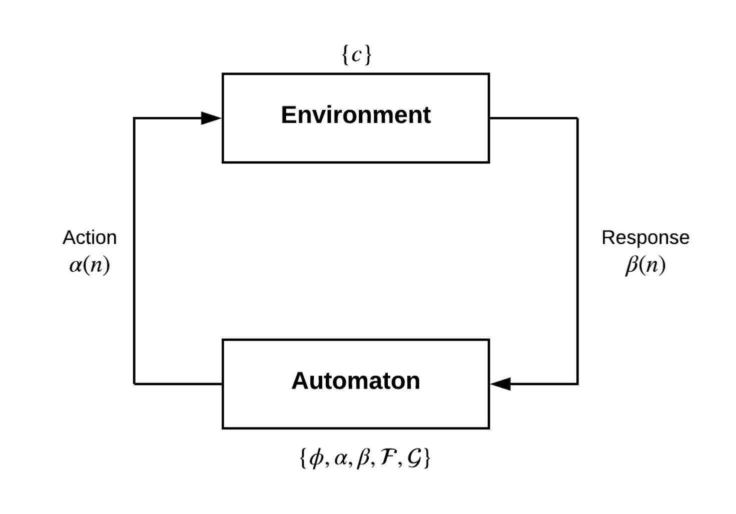


Figure 1: Learning Automata - The Learning Model.

Problem Statement

- LA attempts to minimize the overall cost in NSE over time as:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{r} \mathbb{E}[\beta(n)].$$

- How can we further minimize query accesses for data-structures in NSEs?

Adaptive Data “Sub”-Structures

Move-To-Front (MTF) Rule

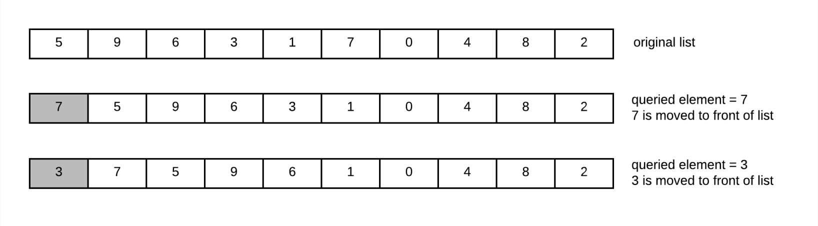


Figure 2: A diagrammatic description of the Move-To-Front (MTF) rule.

The Transposition (TR) Rule

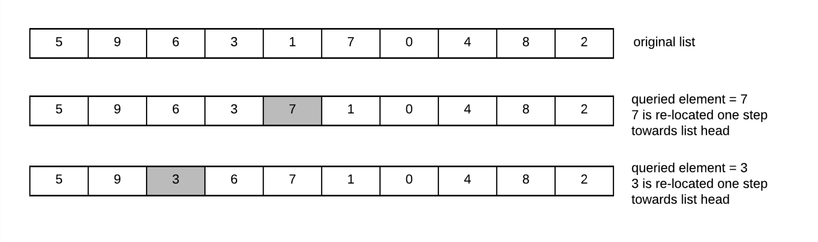


Figure 3: A diagrammatic representation of the Transposition Rule (TR).

The Hierarchical Data “sub”-structure

Singly Linked Lists (SLLs) on Singly Linked Lists.

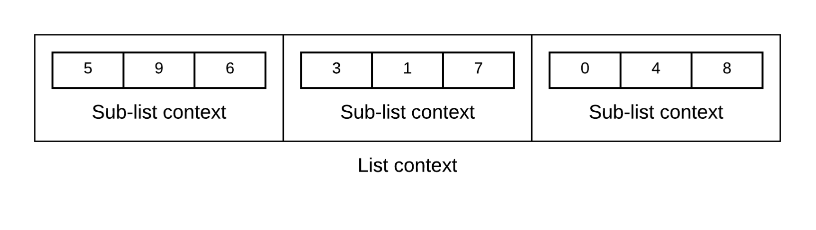


Figure 4: Hierarchical SLLs-on-SLLs

SLLs-on-SLLs Design

- The Design of an Adaptive SLLs-on-SLLs:
  - MTF-MTF.
  - MTF-TR.
  - TR-MTF.
  - TR-TR.
- The static sub-list problem.

MTF-MTF hierarchical adaptive scheme

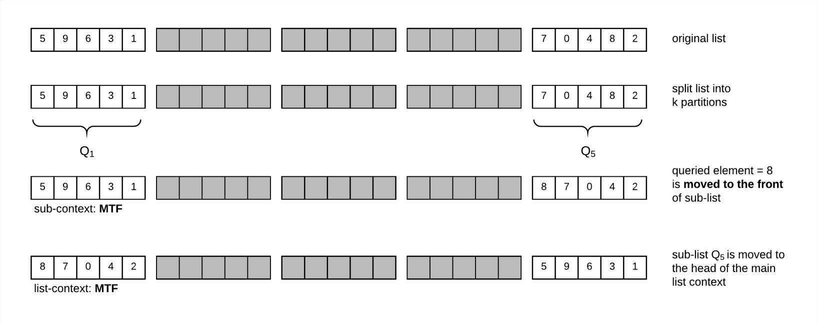


Figure 5: A diagrammatic description of the MTF-MTF.

The static sub-list problem

Scheme	Zipf	Exponential
MTF	43.35	8.68
TR	55.48	10.59
MTF-MTF	59.82	24.78
MTF-TR	59.66	20.47
TR-MTF	61.43	25.70
TR-TR	60.97	21.36

Asymptotic costs in a MSE with $\alpha = 0.9$. Split into 8 sublists.

The Partitioning Problem

Object Partitioning Problem (OPP):

- The partitioning of $\mathcal{W}$ objects into $\mathcal{R}$ groups:
  $\Omega^\star = \{G_1^\star, G_2^\star, \cdots, G_R^\star\}$.
- *NP*-hard problem.
- We would like to find **the most realizable partitioning**.
- Equi-Partitioning Problem (EPP).

The Equi-Partitioning Problem (EPP)

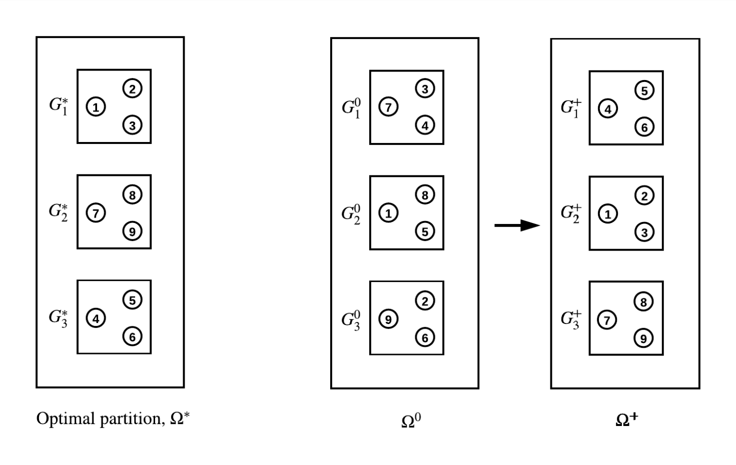
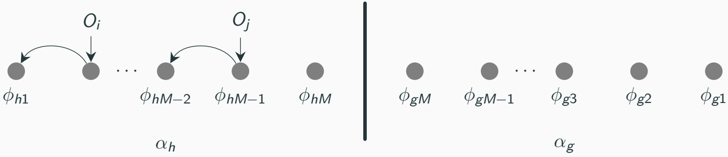


Figure 6: The EPP

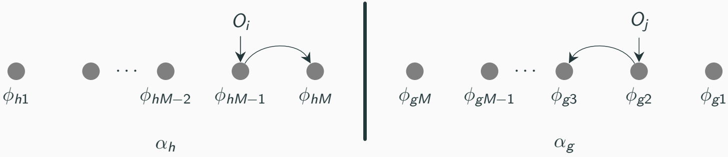
Object Migration Automaton (OMA)

- Faster rate of convergence.
- Simple to implement.
- Superior results.

OMA: Reward and Penalty internal cases



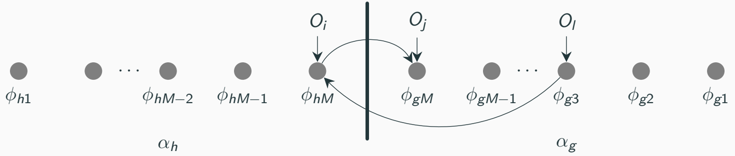
(a) On reward: Move the accessed abstract objects towards the internal states. (Case 1)



(b) On penalty: If both abstract objects are at their boundary states (Case 2).

Figure 7: Reward and Penalty internal cases: $\alpha_{h,g}$ represent the actions.

OMA: Penalty for boundary cases



(c) On penalty: If both abstract objects are in the boundary states (Case 3).

Figure 8: Penalty for boundary cases: $\alpha_{h,g}$ represent the actions.

Objectives

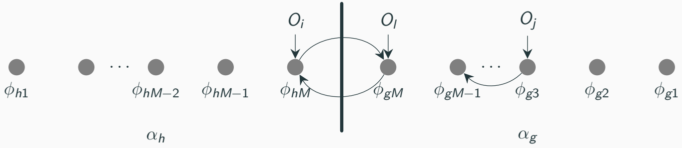
- Design enhanced hierarchical SLL data “sub”-structure using:
  - EOMA.
  - PEOMA.
  - TPEOMA.
- Experiment setup.
  - 128 elements, W, split into $R = \{2, 4, 8, 16, 32, 64\}$ partitions.
  - 10 experiments with 300,000 simulations each.
- Analysis of the new designs.
  - Asymptotic & Amortized costs in MSEs and PSEs.
  - Rate of Convergence with time.

EOMA-Augmented Hierarchical SLLs

EOMA-Augmented Hierarchical SLLs

The Enhancements to the OMA include:

- Solving the Deadlock scenario.
- Redefined Internal States for convergence criteria.



(d) On penalty: A new case for the case when only one object is in the boundary state (Case 4).

Figure 9: Penalty for boundary cases: $\alpha_{h,g}$ represent the actions.

EOMA-Augmented Hierarchical SLLs

- EOMA hierarchical SLLs include:
  - MTF-MTF-EOMA.
  - MTF-TR-EOMA.
  - TR-MTF-EOMA.
  - TR-TR-EOMA.

Performance: EOMA-Augmented SLLs-on-SLLs

Scheme	Zipf	Exponential
MTF	43.35	8.72
TR	55.44	10.52
MTF-MTF-EOMA	19.14	12.34
MTF-TR-EOMA	27.80	16.89
TR-MTF-EOMA	18.84	12.87
TR-TR-EOMA	27.55	17.17

Asymptotic cost in a **MSE** with $\alpha = 0.9$. Split into 8 sublists.

Scheme	Zipf	Exponential
MTF	49.64	8.46
TR	55.65	11.18
MTF-MTF-EOMA	14.63	8.59
MTF-TR-EOMA	25.82	13.88
TR-MTF-EOMA	14.63	8.92
TR-TR-EOMA	25.58	13.70
MTF-MTF-EOMA-P	7.16	6.14
MTF-MTF-EOMA-UP	7.69	8.90

Asymptotic cost in a **PSE** with $T = 30$. Split into 8 sublists.

Performance: EOMA-Augmented SLLs-on-SLLs

- The hierarchical schemes were mostly superior.
- The MTF and TR advantage with the L-shaped distributions.
- MTF rule as outer-list context superior to TR.
- Periodic variations showed superior performances.

Rate of convergence in the MSE

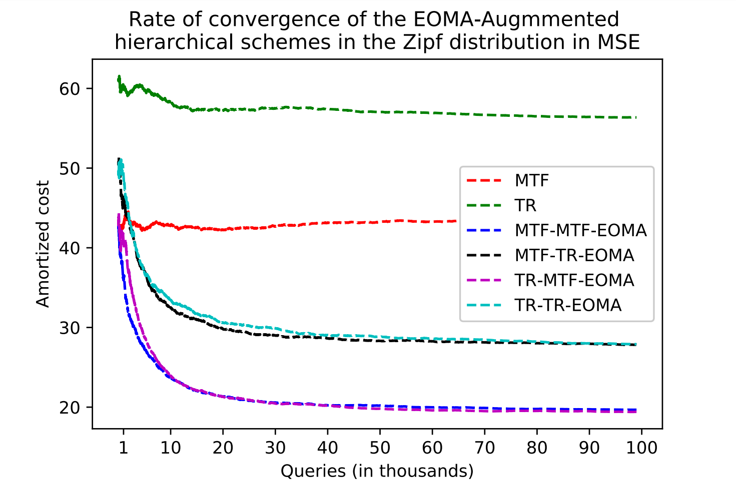


Figure 10: Rate of convergence of the first 100,000 queries for the stand-alone and the EOMA-augmented hierarchical schemes in a MSE.

PEOMA-Augmented Hierarchical SLLs

The Pursuit Concept

- Maximum Likelihood (ML)-based estimates.
- Filter divergent queries from the Environment.

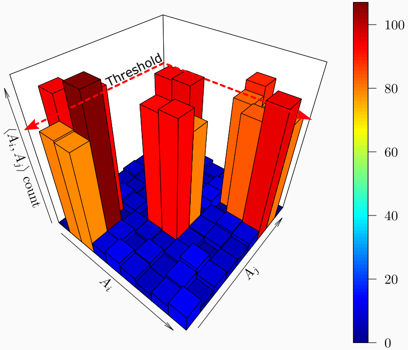


Figure 11: The pursuit concept.

PEOMA-Augmented Hierarchical SLLs

- Hierarchical SLLs augmented with the PEOMA include:
  - MTF-MTF-PEOMA.
  - MTF-TR-PEOMA.
  - TR-MTF-PEOMA.
  - TR-TR-PEOMA.

Performance: PEOMA-Augmented SLLs-on-SLLs

Scheme	Zipf	Exponential
MTF	43.35	8.72
TR	55.44	10.52
MTF-MTF-PEOMA	5.80	2.45
MTF-TR-PEOMA	5.35	2.97
TR-MTF-PEOMA	4.67	2.88
TR-TR-PEOMA	5.07	2.99

Asymptotic cost in a **MSE** with $\alpha = 0.9$. Split into 8 sublists.

Scheme	Zipf	Exponential
MTF	49.64	8.46
TR	55.65	11.18
MTF-MTF-PEOMA	15.19	4.10
MTF-TR-PEOMA	27.30	5.56
TR-MTF-PEOMA	14.90	4.32
TR-TR-PEOMA	27.00	5.56
MTF-MTF-PEOMA-P	7.13	6.10
MTF-MTF-PEOMA-UP	7.71	5.42

Asymptotic cost in a **PSE** with $T = 30$. Split into 8 sublists.

Performance: PEOMA-Augmented SLLs-on-SLLs

- An order of magnitude superior performances in NSEs.
- Thrives even in L-shaped distributions.

Rate of convergence in the MSE

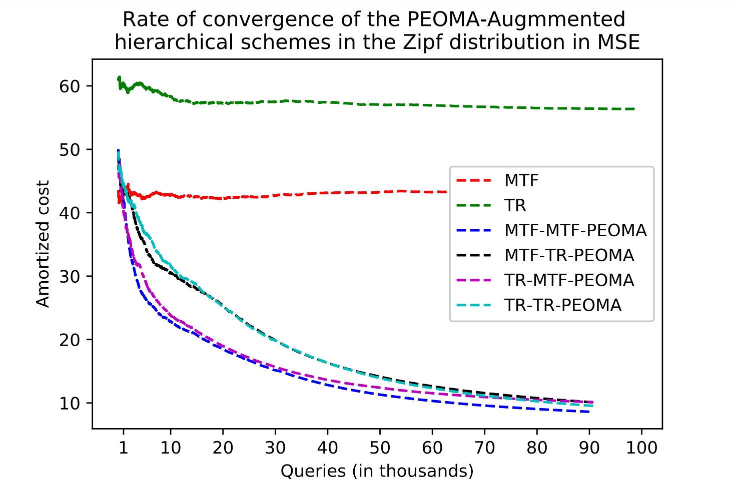


Figure 12: Rate of convergence of the first 100,000 queries for the stand-alone and the PEOMA-augmented hierarchical schemes in a MSE.

TPEOMA-Augmented Hierarchical SLLs

The Transitivity Relation

- Maximum Likelihood (ML)-based estimates.
- Infer good query pairs from non-accessed elements in the transitivity relation.

We need to verify:

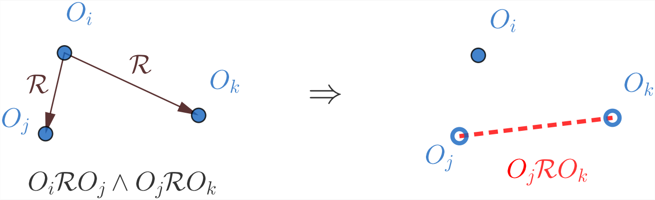


Figure 13: The transitivity relation.

TPEOMA-Augmented Hierarchical SLLs

- Hierarchical SLLs augmented with the TPEOMA include:.
  - MTF-MTF-TPEOMA.
  - MTF-TR-TPEOMA.
  - TR-MTF-TPEOMA.
  - TR-TR-TPEOMA.

Performance: TPEOMA-Augmented SLLs-on-SLLs

Scheme	Zipf	Exponential
MTF	43.35	8.72
TR	55.44	10.52
MTF-MTF-TPEOMA	13.45	7.01
MTF-TR-TPEOMA	11.51	6.99
TR-MTF-TPEOMA	12.50	7.50
TR-TR-TPEOMA	14.80	7.64

Asymptotic cost in a **MSE** with $\alpha = 0.9$. Split into 8 sublists.

Scheme	Zipf	Exponential
MTF	49.64	8.46
TR	55.65	11.18
MTF-MTF-TPEOMA	24.16	7.24
MTF-TR-TPEOMA	26.94	7.15
TR-MTF-TPEOMA	25.16	7.44
TR-TR-TPEOMA	27.99	7.32
MTF-MTF-TPEOMA-P	25.41	11.08
MTF-MTF-TPEOMA-UP	53.30	14.10

Asymptotic cost in a **PSE** with $T = 30$. Split into 8 sublists.

Performance: TPEOMA-Augmented SLLs-on-SLLs

- Good performances when the outer-list context rule was MTF.
- However, not of the quality of results we expected.
- Periodic variations performances were inferior.
- TPEOMA-Augmented SLLs not recommended for PSEs.

Rate of convergence in the MSE

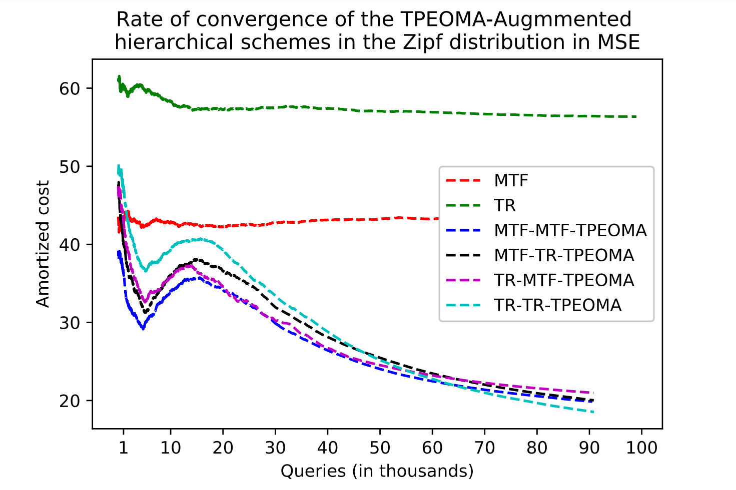


Figure 14: Rate of convergence of the first 100,000 queries for the stand-alone and the TPEOMA-augmented hierarchical schemes in a MSE.

Summary/ Conclusion

Summary

- ADS minimizes asymptotic search cost in NSEs.
- LOLs using state-of-the-art OMA reinforcement schemes.
- Results showed superior performances.

Conclusion

- The EOMA hierarchical schemes had *mostly* superior results.
- PEOMA hierarchical schemes were an order of magnitude superior.
- The TPEOMA hierarchical schemes are unsuitable for PSEs.
- Periodic variations were superior for EOMA/PEOMA in PSEs.
- The advanced LOL-SLLs are the state-of-the-art ADSs in NSEs.

Future Work

Future Work

- Hierarchical LOLs for:
  - doubly-linked-lists on singly-linked-lists,
  - singly-linked-lists on doubly-linked-list and for
  - doubly-linked-lists on doubly-linked-lists.
- Each of these topics are potential theses in their own rights.
- Package these solutions as a programming library.
- The rigorous formal analysis of these schemes are open.

Any Questions?

Appendices

Ratio of the MTF-MTF-EOMA to the MTF scheme in MSEs

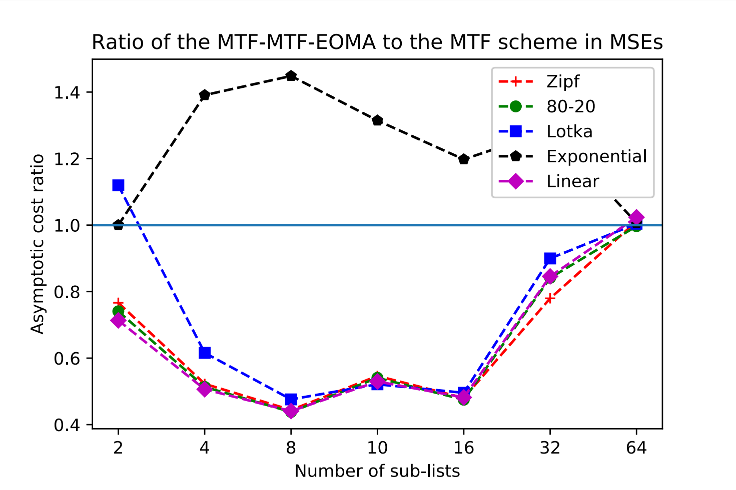


Figure 15: The asymptotic cost ratio of the MTF-MTF-EOMA to the MTF scheme for sub-list variations in the MSE ($\alpha = 0.9$).

Ratio of the TR-TR-EOMA to the MTF scheme in MSEs

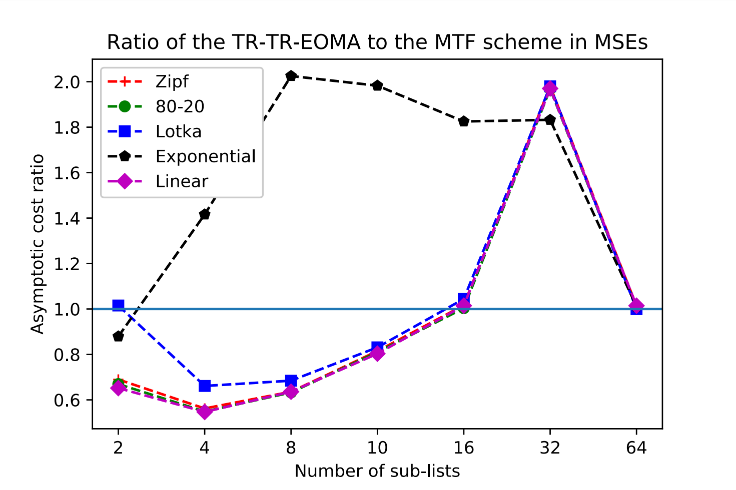


Figure 16: The asymptotic cost ratio of the TR-TR-EOMA to the MTF scheme for sub-list variations in the MSE ($\alpha = 0.9$).

Asymptotic cost with varying dependence degrees

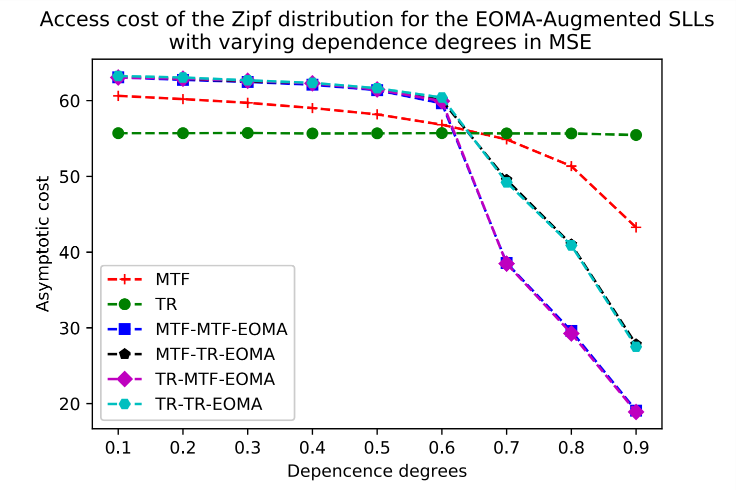


Figure 17: Changes in the asymptotic cost of the stand-alone and hierarchical schemes with EOMA in the MSE.

Ratio of the TR-MTF-EOMA to the MTF scheme in PSEs

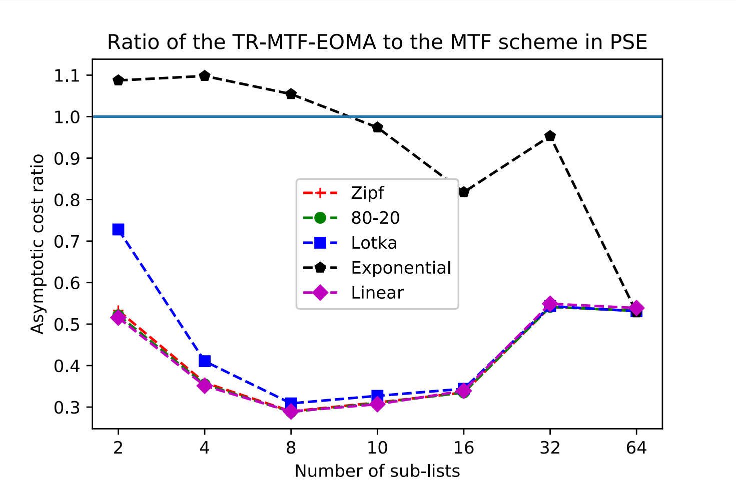


Figure 18: The asymptotic cost ratio of the TR-MTF-EOMA:MTF in PSE with period $T = 30$.

Ratio of the MTF-TR-EOMA to the MTF scheme in PSEs

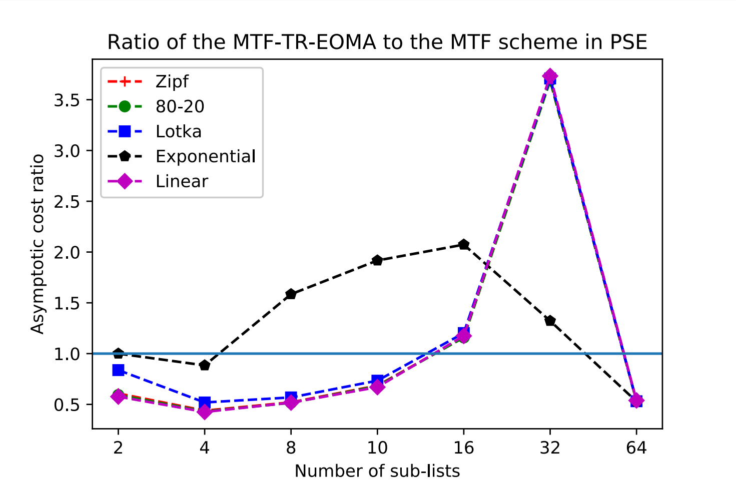


Figure 19: The asymptotic cost ratio of the MTF-TR-EOMA:MTF in PSE with period $T = 30$.

Asymptotic cost with varying periods in the PSE

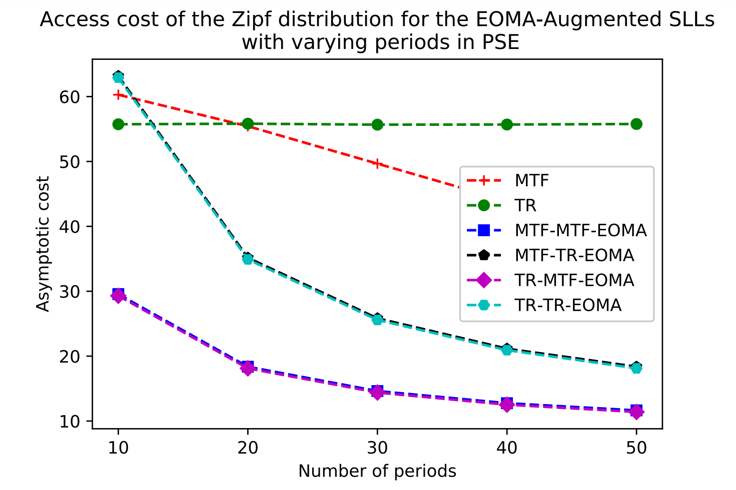


Figure 20: Changes in the asymptotic cost of the stand-alone and hierarchical schemes with EOMA in the PSE.

Asymptotic cost for Periodic variations of MTF-MTF-EOMA

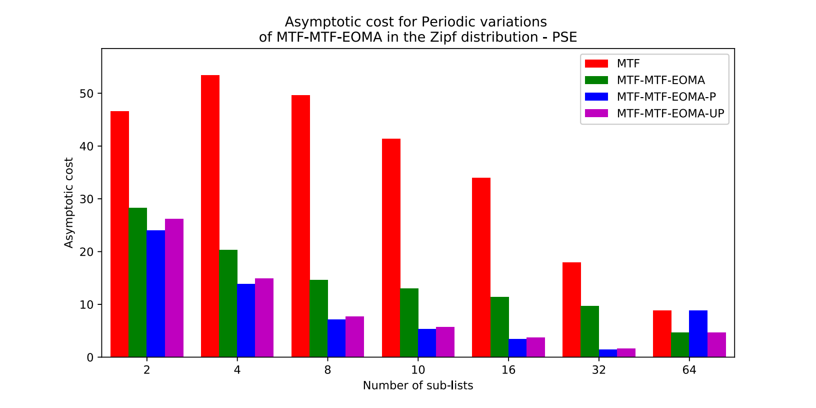


Figure 21: Asymptotic cost for Periodic variations of MTF-MTF-EOMA in the PSE-Zipf distribution with period $T = 30$.

Ratio of the MTF-MTF-PEOMA to the MTF scheme in MSEs

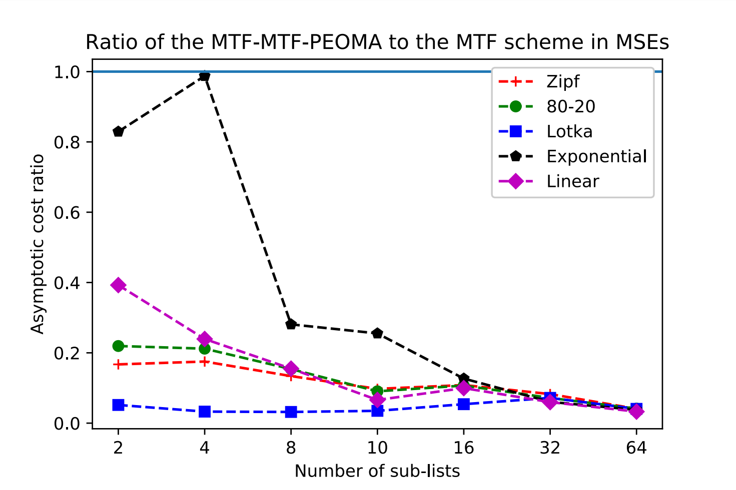


Figure 22: The asymptotic cost ratio of the MTF-MTF-PEOMA:MTF in MSE with $\alpha = 0.9$.

Ratio of the TR-TR-PEOMA to the MTF scheme in MSEs

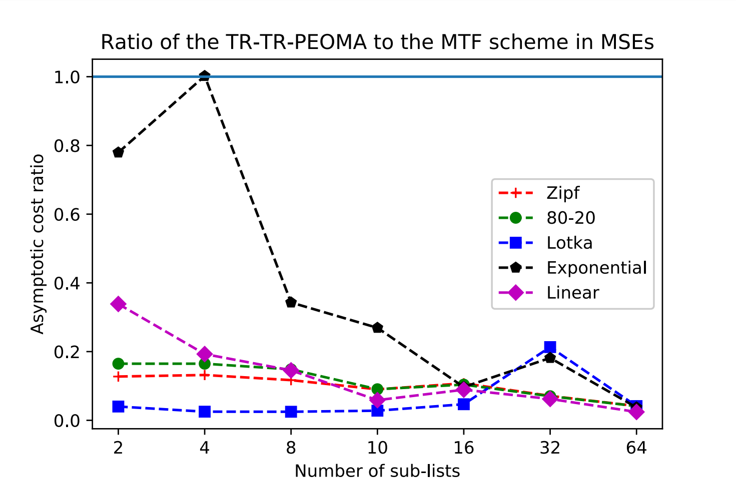


Figure 23: The asymptotic cost ratio of the TR-TR-PEOMA:MTF in MSE with $\alpha = 0.9$.

Asymptotic cost with varying dependence degrees

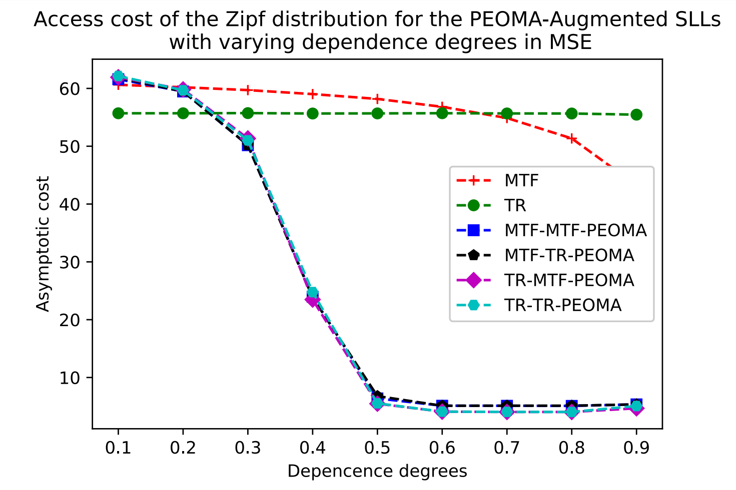


Figure 24: Changes in the asymptotic cost of the stand-alone and hierarchical schemes with PEOMA in the MSE.

Ratio of the TR-MTF-PEOMA to the MTF scheme in PSEs

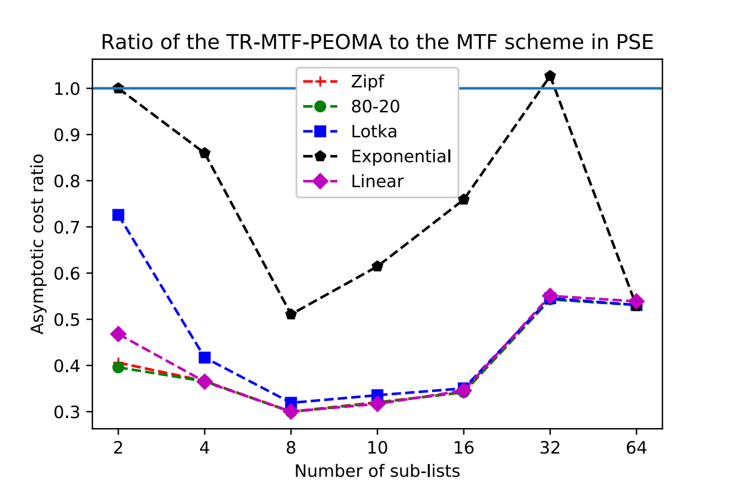


Figure 25: The asymptotic cost ratio of the TR-MTF-PEOMA:MTF in PSE with period $T = 30$.

Ratio of the MTF-TR-PEOMA to the MTF scheme in PSEs

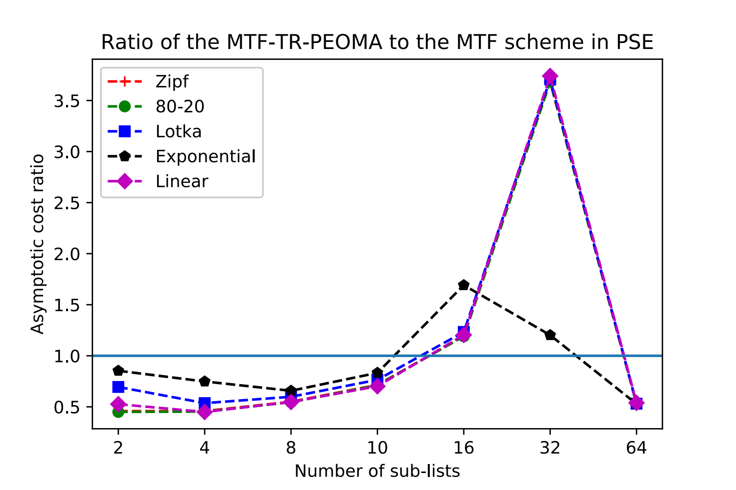


Figure 26: The asymptotic cost ratio of the MTF-TR-PEOMA:MTF in PSE with period $T = 30$.

Asymptotic cost for Periodic variations of TR-MTF-PEOMA

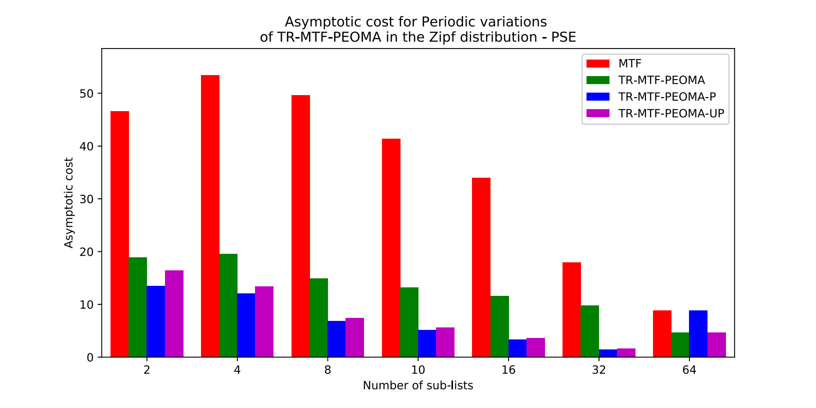


Figure 27: Asymptotic cost for Periodic variations of TR-MTF-PEOMA in the PSE-Zipf distribution with period $T = 30$.

Asymptotic cost for Periodic variations of MTF-TR-PEOMA

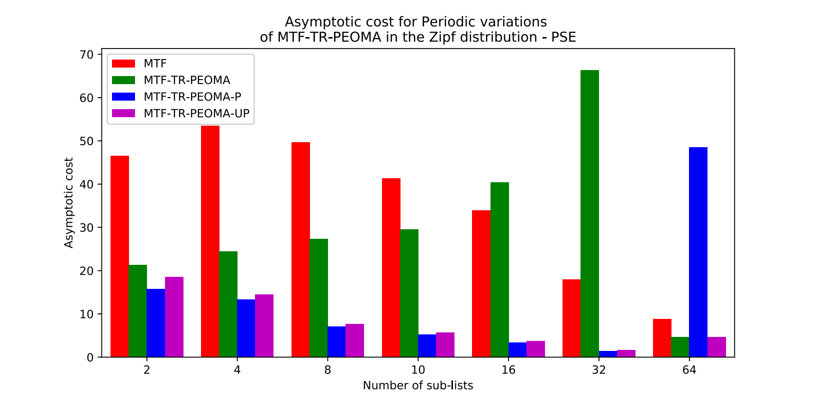


Figure 28: Asymptotic cost for Periodic variations of MTF-TR-PEOMA in the PSE-Zipf distribution with period $T = 30$.

Ratio of the TR-MTF-TPEOMA to the MTF scheme in MSEs

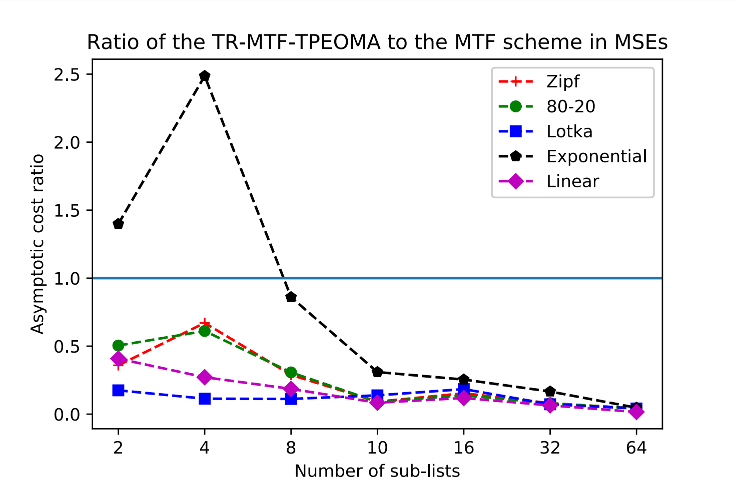


Figure 29: The asymptotic cost ratio of the TR-MTF-TPEOMA:MTF in MSE with $\alpha = 0.9$.

Ratio of the TR-TR-TPEOMA to the MTF scheme in MSEs

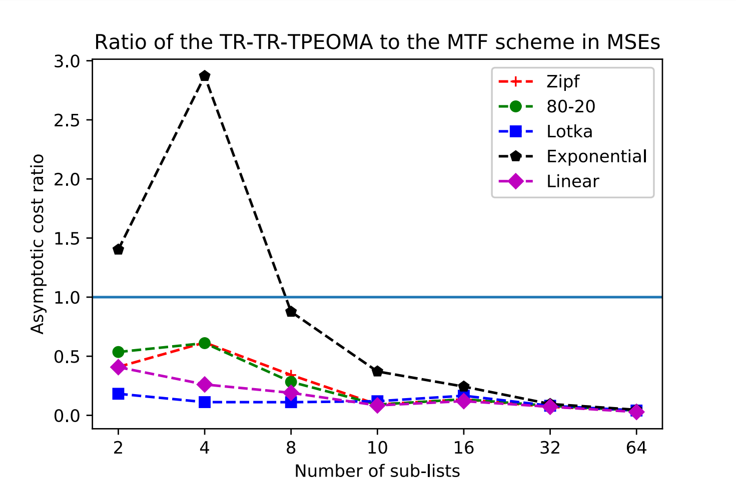


Figure 30: The asymptotic cost ratio of the TR-TR-TPEOMA:MTF in MSE with $\alpha = 0.9$.

Asymptotic cost with varying dependence degrees

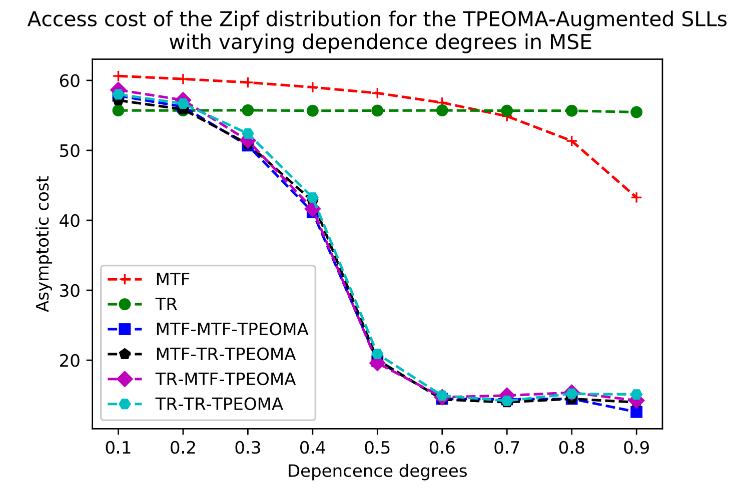


Figure 31: Changes in the asymptotic cost of the stand-alone and hierarchical schemes with TPEOMA in the MSE.

Ratio of the TR-MTF-TPEOMA to the MTF scheme in PSEs

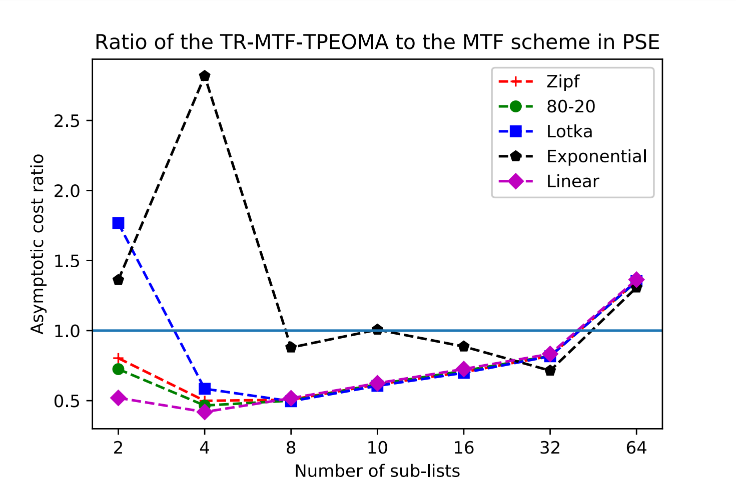


Figure 32: The asymptotic cost ratio of the TR-MTF-TPEOMA:MTF in PSE with period $T = 30$.

Ratio of the MTF-TR-TPEOMA to the MTF scheme in PSEs

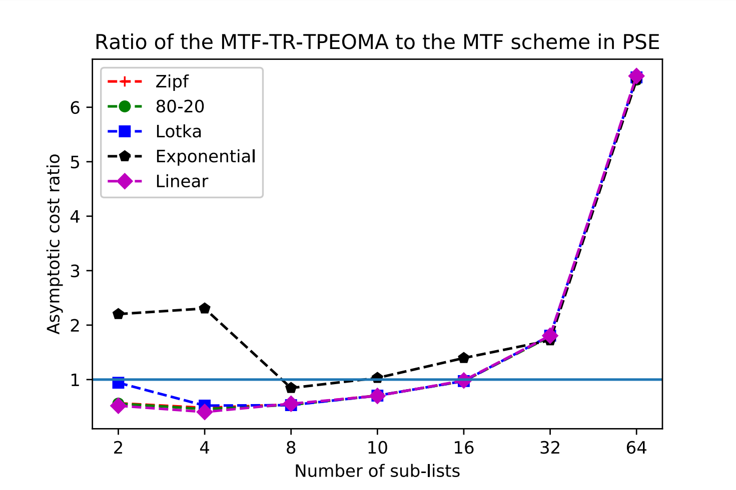


Figure 33: The asymptotic cost ratio of the MTF-TR-TPEOMA:MTF in PSE with period $T = 30$.

Asymptotic cost with varying periods in the PSE

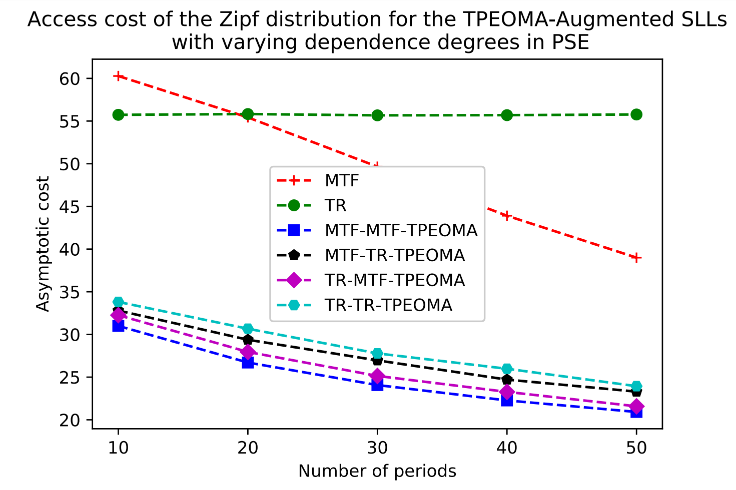


Figure 34: Changes in the asymptotic cost of the stand-alone and hierarchical schemes with TPEOMA in the PSE.

Asymptotic cost for Periodic variations of TR-MTF-TPEOMA

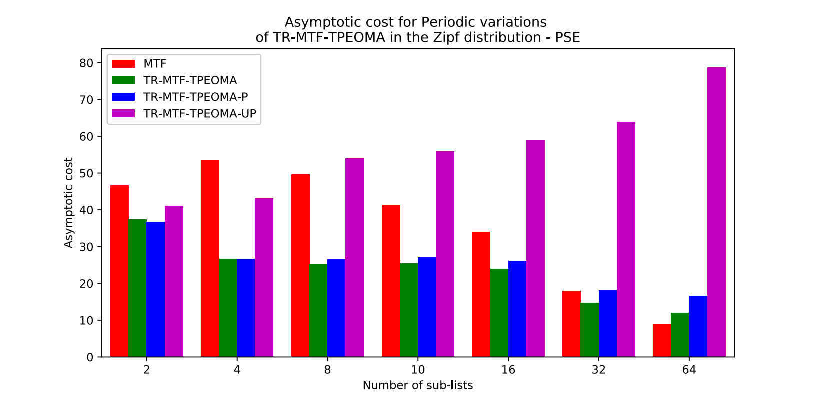


Figure 35: Asymptotic cost for Periodic variations of TR-MTF-TPEOMA in the PSE-Zipf distribution with period $T = 30$.