IJCAI
JEJU 2024

FlagVNE:
A Flexible and Generalizable Reinforcement Learning Framework for Network Resource Allocation

AI for Networking

Tianfu Wang, Qilin Fan, Chao Wang, Long Yang,
Leilei Ding, Nicholas Jing Yuan, Hui Xiong

University of Science and Technology of China (USTC)

Chongqing University

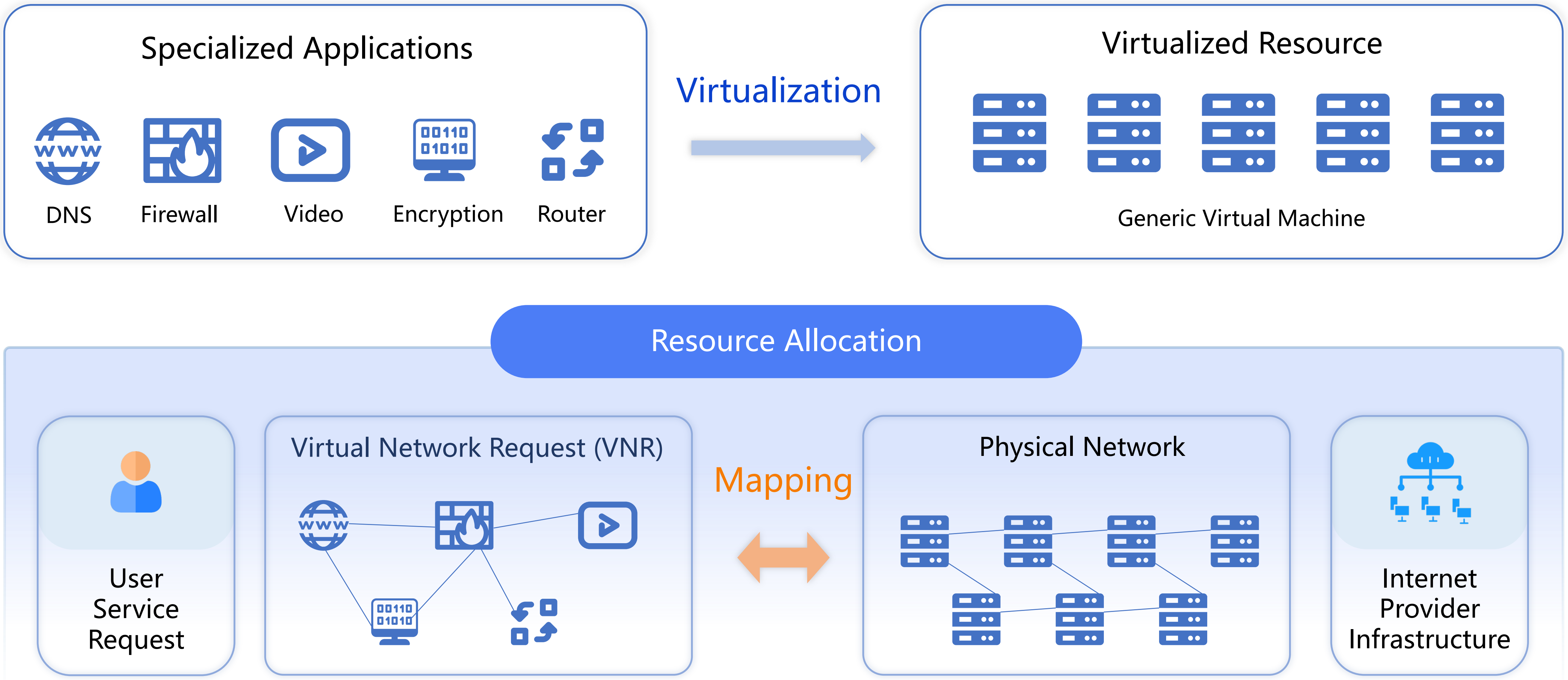
Peking University

Microsoft Inc.

The Hong Kong University of Science and Technology (Guangzhou)

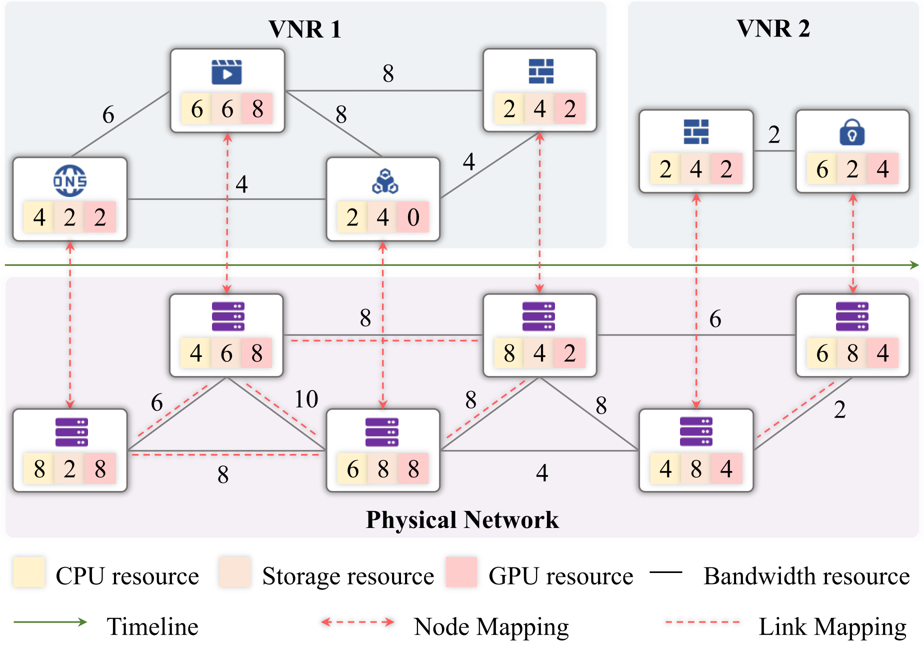
Background | **Network Virtualization**

Enable the dynamic management of Internet architecture (e.g., 5G networks and cloud computing)



Background | **Virtual Network Embedding Problem**

Maps VNRs to physical network with minimal resource cost while satisfying various constraints



Node Mapping

Assign each virtual node to a physical node with sufficient resources

- One-to-one Placement
- Node Resource Availability

Link Mapping

Finding connective physical paths for virtual links between nodes

- Path Connectivity
- Link Resource Availability

Problem Characteristics

Combinatorial Explosion

NP-hard Combinatorial Optimization Problem

The extensive solution space requires comprehensive exploration

Differentiated Demands

Specific requirements of user service are diverse

VNRs have varied sizes and resource requirements

Background | **Literature Review**

Reinforcement learning (RL) has shown promising potential for the VNE problem

Traditional Methods

Exact Methods

Provide optimal solutions

impractical for real-time scenarios
high computational costs.

Heuristics

offer faster solutions

often require manual design
are scenario-specific.

Machine Learning for VNE

Supervised Learning

Rely on labeled data

Reinforcement Learning

Solution construction process of VNR

▼

Markov decision processes (MDP)

Automatically learning of effective heuristics through interactions with the environment

Background | **Motivations & Challenges Inspired by Preliminary Study**

Motivation A: Flexibility of Action Space

Existing Methods

They employ a unidirectional action design, i.e., assuming that decision sequence of virtual nodes is predetermined

Preliminary Study

Figure 4 reveals that varying the decision sequences of virtual nodes significantly impacts performance

Intuitive Direction

Achieve a joint selection of both physical and virtual nodes to enhance the flexibility of exploration and exploitation

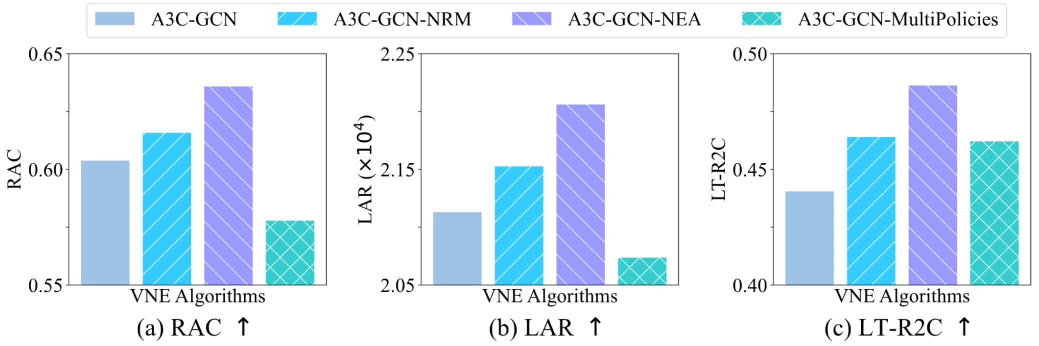


Figure 4: Comparative performance of A3C-GCN variants on three metrics: Impact of decision sequence and size-specific policies on VNE. (We conduct experiments using WX100 as the physical network, with a VNR arrival rate of 0.18. All other settings remained consistent with those described in Section 5.)

Latent Challenges

The difficulty of variable action prob distribution generation

The training efficiency issue caused by large action space

Background | **Motivations & Challenges Inspired by Preliminary Study**

Motivation B. Generalization of Solving Policy

Existing Methods

They typically use a one-size-fits-all policy to tackle VNRs of varying sizes, leading to generalization issues

Preliminary Study

Figure 5 reveals that some size-specific policies are superior or to single-policy, while some are inferior.

Intuitive Direction

Train a set of sub-policies directly to handle VNRs of different sizes from scratch

Latent Challenges

Specific policies trained from scratch encounter local optima

How to quickly adapt to handle previously unseen VNR sizes

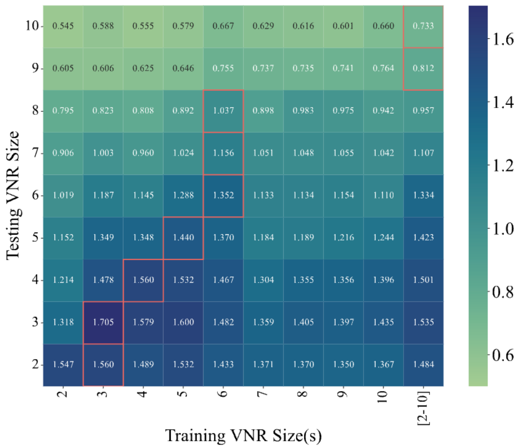


Figure 5: Average returns of the one-fits-all-size policy and each size-specific policy on all testing VNR sizes. The red boxes indicate the best performance results for test sizes. In the horizontal axis, [2-10] indicates a well-trained A3C-GCN policy while a single number represents a size-specific policy derived from well-trained A3C-GCN-MultiPolicy. (We use WX100 as the physical network and all training settings are the same as those mentioned in Section 5. For testing data of each VNR size, to exclude network system dynamics for a fairer comparison, we randomly generated 1000 static instances, including VNR and physical networks, as the benchmark. The performance metric is defined as the average episode return over 1000 instances.)

Method | **Overview of FlagVNE Framework**

A **Flexible** and **Generalizable** Reinforcement Learning Framework for Solve VNE problem

Flexibility

Efficiency

Generalizability

A. Bidirectional Action-based MDP

Joint selection of virtual & physical nodes

Enhance the flexibility of agent exploration and exploitation

B. Hierarchical Policy Architecture

High-level ordering policy

Select the appropriate virtual node for the low-level placement.

Low-level placement policy

Identify a suitable physical node for placing to-be-placed virtual node

Distribution Size: $|N^v| \times |N^p| \rightarrow |N^v| + |N^p|$

Adaptively generate action prob dists and ensure high training efficiency.

C. Generalizable Training Method

Meta-RL for VNE

Formulate varying VNRs as distinct MDPs based their size.

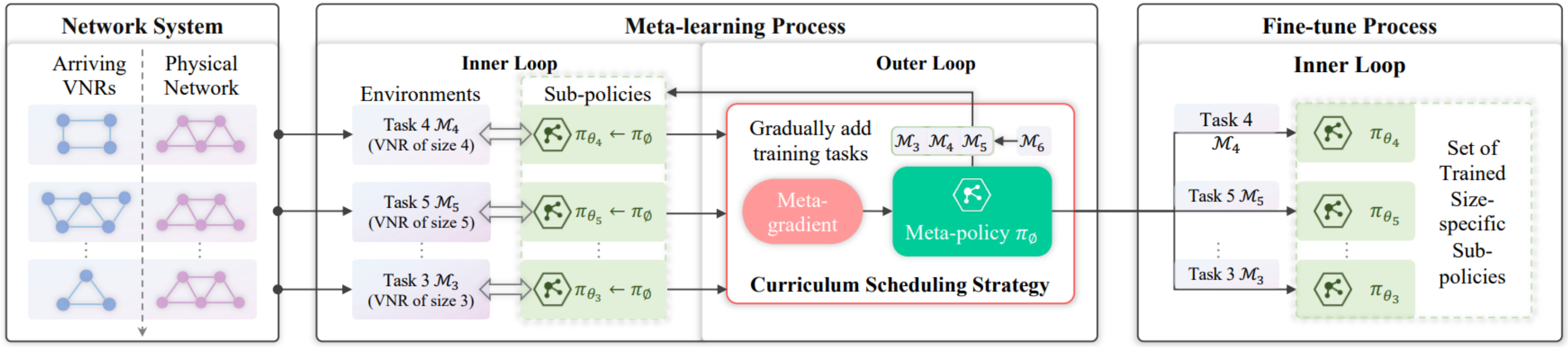
Adopt model-agnostic meta-learning (MAML) as the basic training method

Curriculum Scheduling Strategy

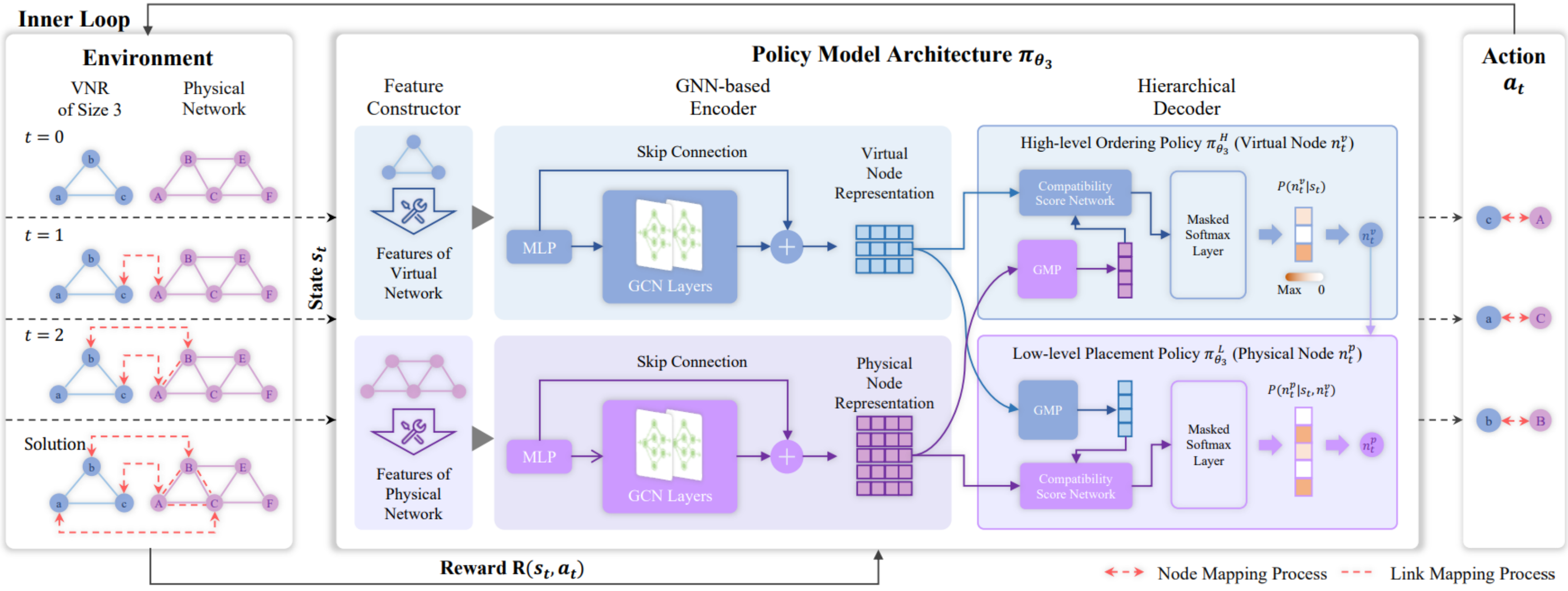
Training large-size policies from scratch adversely impacting meta-learning

Gradually increase the complexity of tasks during RL training

Effectively obtain refined solving policies for each VNR size



(a) Generalizable training method for obtaining multiple size-specific sub-policies



(b) Bidirectional action-based MDP modeling within each inner loop (sub-policy π_{θ_3} for example)

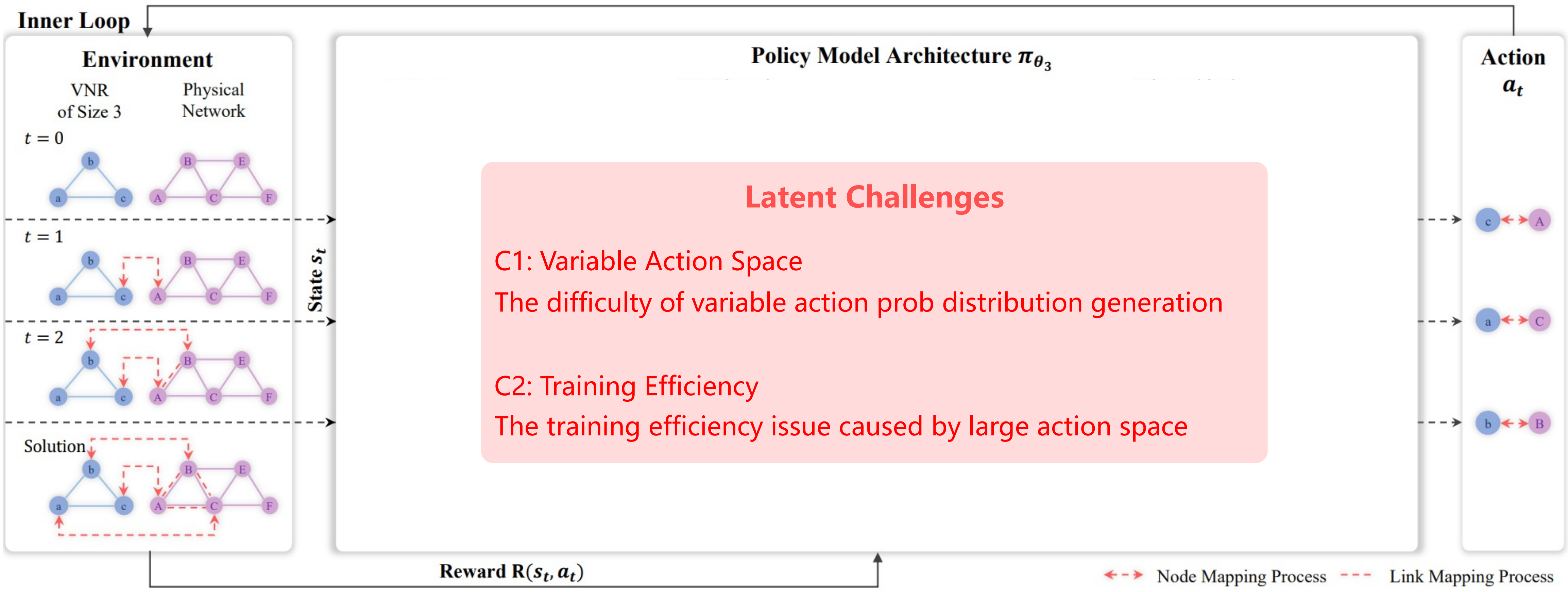
Method | **A. Bidirectional Action-based MDP**

Allows the joint selection of virtual nodes and physical nodes.

Enhances the flexibility and comprehensiveness of exploration.

Increased Flexibility

Improved Solution Quality



Method | B. Hierarchical Policy Architecture

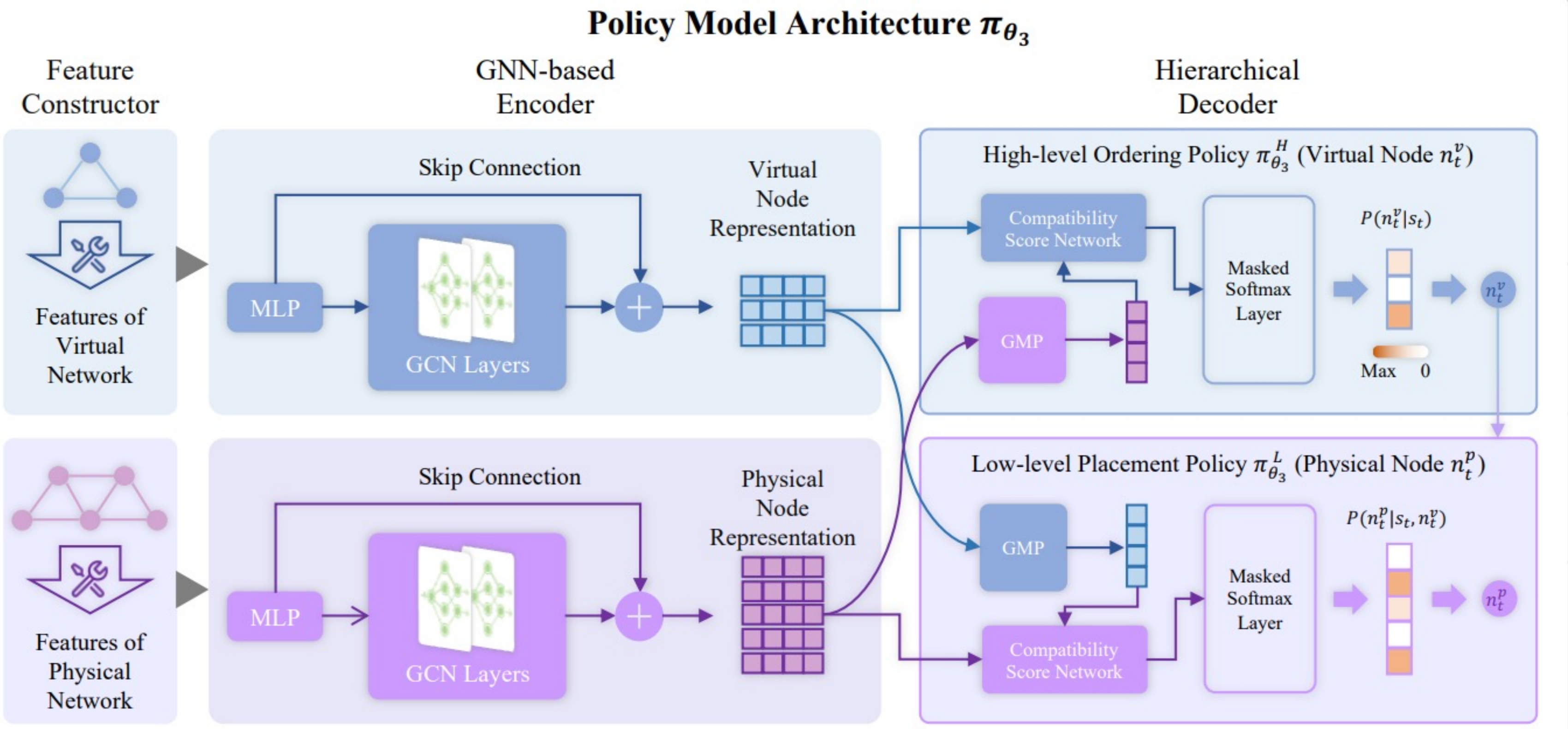
High-level Ordering Policy: Selects the virtual node to be placed based on compatibility scoring.

Low-level Placement Policy: Chooses the physical node to host the selected virtual node.

Hierarchical structure reduces the complexity of the distribution size: $|N^v| \times |N^p| \rightarrow |N^v| + |N^p|$

Adaptive Action Probability Distribution

Ensure High training Efficiency



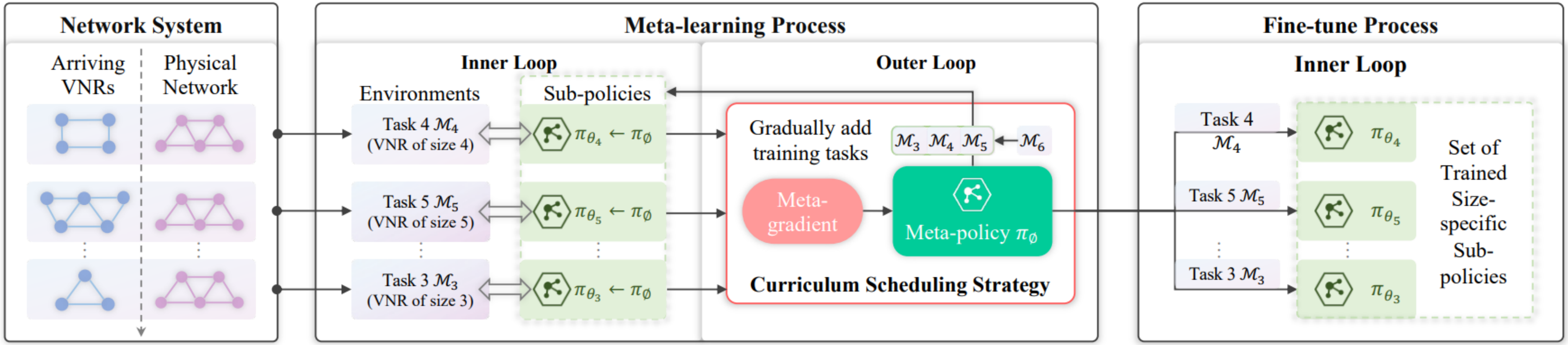
Method | C. Generalizable Training Method

Meta-RL Training Approach with Curriculum Scheduling Strategy

Improved Generalizability

Efficient Training

- **Model-Agnostic Meta-Learning (MAML)**
  - Treats VNRs of different sizes as distinct tasks.
  - Obtain a set of size-specific policies and facilitates fast adaptation to new tasks with limited training samples.
- **Progressive Task Inclusion**
  - Gradually incorporates larger VNRs into the training process.
  - Alleviates suboptimal convergence by ensuring high-quality initializations for large VNR tasks.



(a) Generalizable training method for obtaining multiple size-specific sub-policies

Experiment | **Setup & Implementations**

Simulation Environment

Topologies of Physical Network

- GEANT: 40 nodes, 61 links.
- WX100: 100 nodes, 500 links.

VNR Generation:

- Quantity: 1000 VNRs per simulation run.
- Size: VNRs with 2 to 10 nodes.
- Resource Demands:
    - Node resources within [0, 20] units
    - Link bandwidth within [0, 50] units.
- Lifetime: Exponentially distributed with an average of 500 time units.
- Arrival Rate: Follows a Poisson process, varied to simulate different traffic throughputs.

Training Process

- **Initial Meta-Learning**: Conducted in the first 20 simulations.
- **Fine-Tuning**: Conducted in the subsequent 10 simulations.

Baselines

Heuristic Methods

- NRM-VNE
- NEA-VNE
- PSO-VNE

RL-based Methods

- MCTS-VNE
- PG-CNN
- A3C-GCN
- DDPG-Attention

Metrics

Request Acceptance Rate **(RAC)**

Long-term Average Revenue **(LAR)**

Long-term Revenue-to-Cost **(LT-R2C)**

Experiment | **Overall Performance**

Both GEANT and WX100 Topology

FlagVNE demonstrates exceptional performance in terms of request acceptance, revenue generation, and cost-effectiveness across different network scenarios.

The improvements are more pronounced in environments with higher resource competition, underscoring the framework's robustness and efficiency.

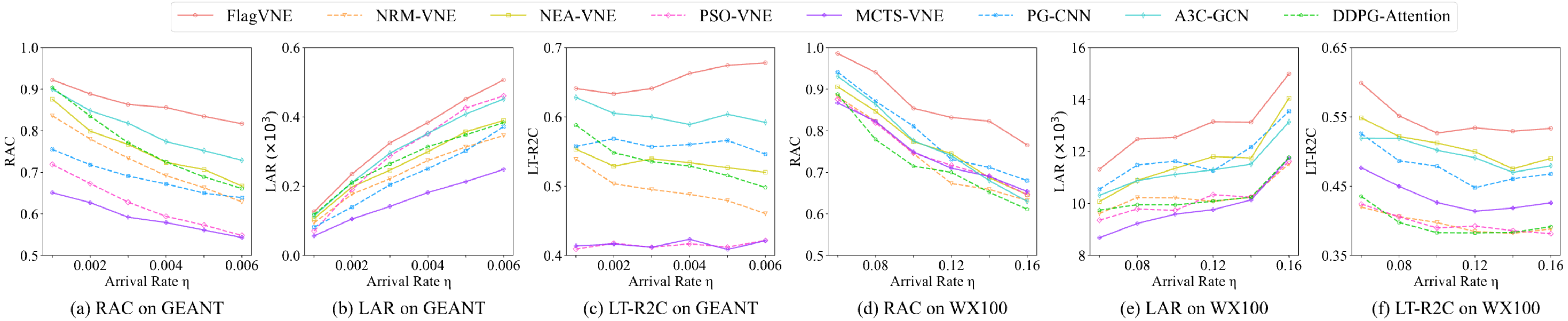


Figure 3: Experimental results in traffic throughput test.

Experiment | **Ablation Study**

To verify the effectiveness of each component in the FlagVNE framework by comparing variations.

Variations of FlagVNE

FlagVNE-UniActionNEA

Replaces the bidirectional action with unidirectional action.

Uses Node Essentiality Assessment (NEA) for decision sequence.

FlagVNE-MetaFree-SinglePolicy

Trains a single general policy without the Meta-RL approach.

FlagVNE-MetaFree-MultiPolicy

Directly trains multiple policies from scratch without Meta-RL.

FlagVNE-MetaPolicy

Uses only the meta-policy for handling VNRs of all sizes.

FlagVNE-NoCurriculum

Discards the curriculum scheduling strategy during the meta-learning process.

	GEANT			WX100		
	RAC ↑	LAR ↑	LT-R2C ↑	RAC ↑	LAR ↑	LT-R2C ↑
FlagVNE-UniActionNEA	0.781	475.335	0.637	0.724	14334.671	0.493
FlagVNE-MetaFree-SinglePolicy	0.758	472.455	0.614	0.712	14170.514	0.501
FlagVNE-MetaFree-MultiPolicy	0.746	435.502	0.593	0.685	14069.938	0.472
FlagVNE-MetaPolicy	0.773	478.646	0.634	0.717	14292.962	0.485
FlagVNE-NoCurriculum	0.787	485.267	0.643	0.708	14144.234	0.509
FlagVNE	**0.804**	**499.303**	**0.668**	**0.754**	**14769.080**	**0.526**

Table 1: Results on ablation study. ($\eta = 0.006$ on GEANT and $\eta = 0.18$ on WX100).

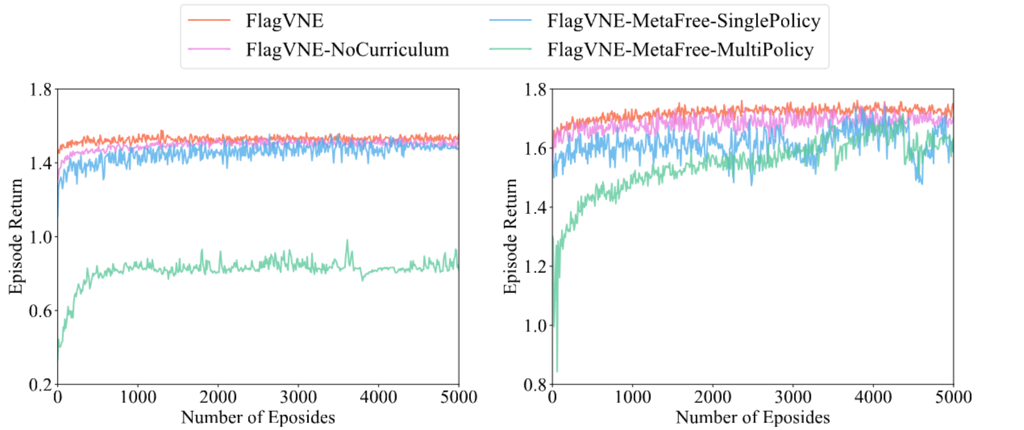


Figure 7: Learning curves for the unseen size. (There are 100 VNRs in one simulation and each VNR's size is set to 12).

Experiment | **Additional Evaluation**

Scalability Validation

Large-scale Topology: 500 nodes and 1000 links..

Algorithm	RAC ↑	LAR ($\times 10^6$) ↑	LT-R2C ↑
NRM-VNE	0.631	0.710772	0.507
NEA-VNE	0.857	1.186615	0.690
PSO-VNE	0.805	1.042604	0.537
MCTS-VNE	0.782	0.968175	0.563
PG-CNN	0.851	1.046523	0.548
A3C-GCN	0.869	1.147116	0.715
DDPG-Attention	0.796	1.013670	0.617
FlagVNE	**0.932**	**1.347162**	**0.744**

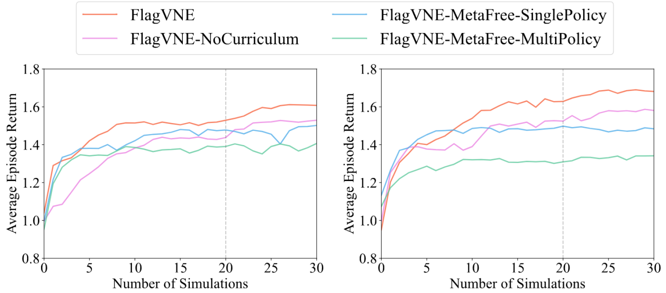
Running Time Test

	Average Running Time (s) ↓	
	GEANT	WX100
NRM-VNE	10.079	28.079
NEA-VNE	31.011	238.403
PSO-VNE	1330.706	1516.340
MCTS-VNE	240.195	679.007
PG-CNN	75.259	203.965
A3C-GCN	47.079	204.073
DDPG-Attention	81.713	164.355
FlagVNE	84.987	239.251

* The average simulation time (seconds) over various η

Table 2: Average running time in traffic throughput test.

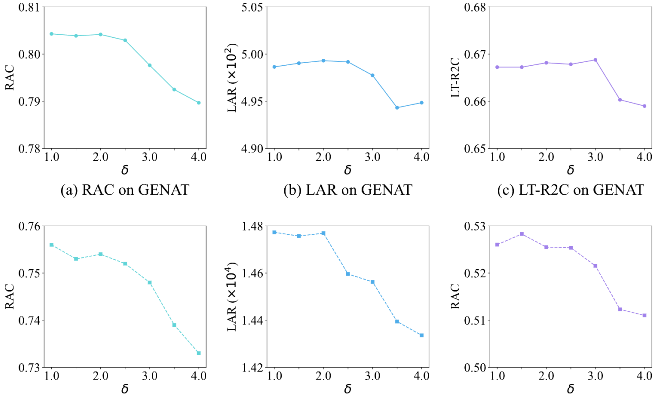
Convergence Analysis



(a) Learning Curve on GEANT (b) Learning Curve on WX100

Figure 6: Learning curves over simulations. Within each simulation, there are 1000 VNRs, i.e., 1000 episodes and we average their returns. ($\eta = 0.001$ on GEANT and $\eta = 0.06$ on WX100.)

Hyperparameter Sensitivity



(a) RAC on GENAT (b) LAR on GENAT (c) LT-R2C on GENAT

(e) RAC on WX100 (f) LAR on WX100 (g) RAC on WX100

Figure 8: The impact of δ on FlagVNE's performance. ($\eta = 0.001$ on GEANT and $\eta = 0.06$ on WX100.)

Conclusion | **FlagVNE Framework**

Preliminary Study

Issue A: Searchability - the unidirectional action design

Issue A: generalizability - one-size-fits-all training strategy

FlagVNE Framework

A bidirectional action-based MDP model

- Jointly select of virtual and physical nodes
- Superior searchability and proven theoretically

A hierarchical decoder with a bilevel policy

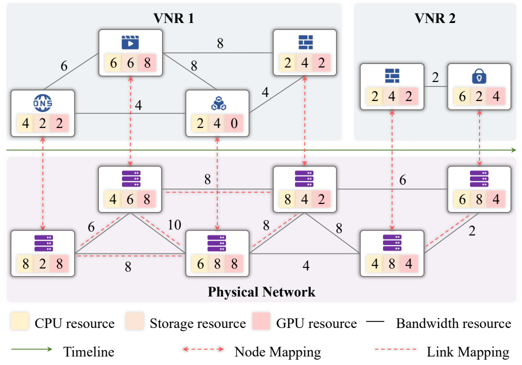
- ensure adaptive action prob dist generation
- ensure high training efficiency

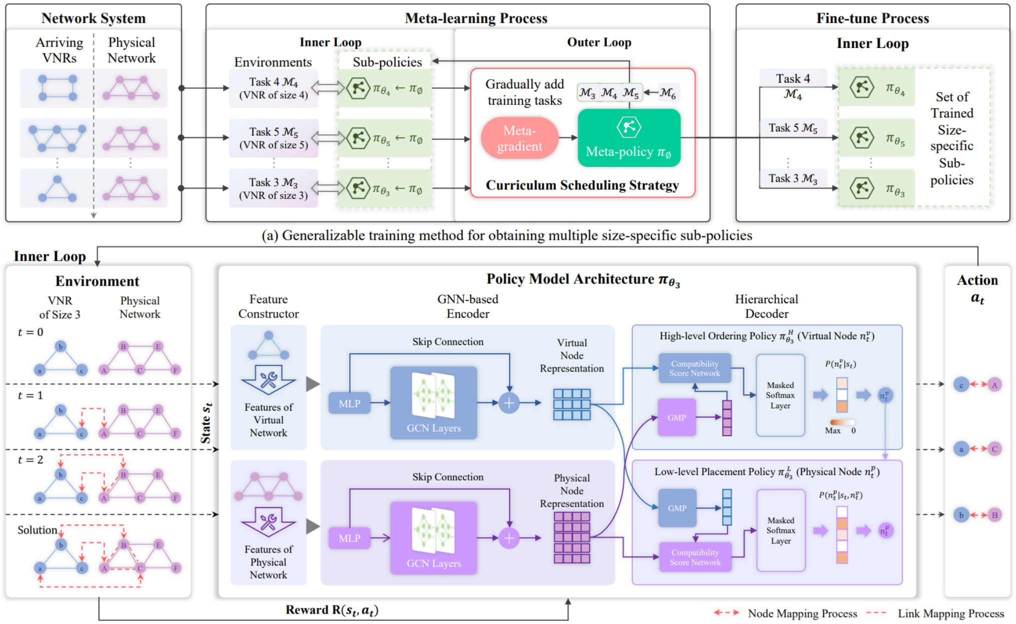
A meta RL-based training method

- efficient obtain multiple size-specific policies
- quick adaptation to new sizes

A curriculum scheduling strategy

- gradually incorporates larger VNRs
- alleviate suboptimal convergence





(a) Generalizable training method for obtaining multiple size-specific sub-policies

(b) Bidirectional action-based MDP modeling within each inner loop (sub-policy π_{θ_3} for example)

Extensive Experiments

IJCAI JEJU 2024

Thanks!

AI for Networking

FlagVNE: A Flexible and Generalizable Reinforcement Learning Framework for Network Resource Allocation

Tianfu Wang, Qilin Fan, Chao Wang, Long Yang, Leilei Ding, Nicholas Jing Yuan, Hui Xiong