Nüwa: Efficient Generative Control Plane for AI Network Simulation

Wenkai Li[1], Ran Shu[2], Peng Zhang[1], Yongqiang Xiong[2]

[1]Xi'an Jiaotong University, [2]Microsoft Research

Why Do We Need Simulation?

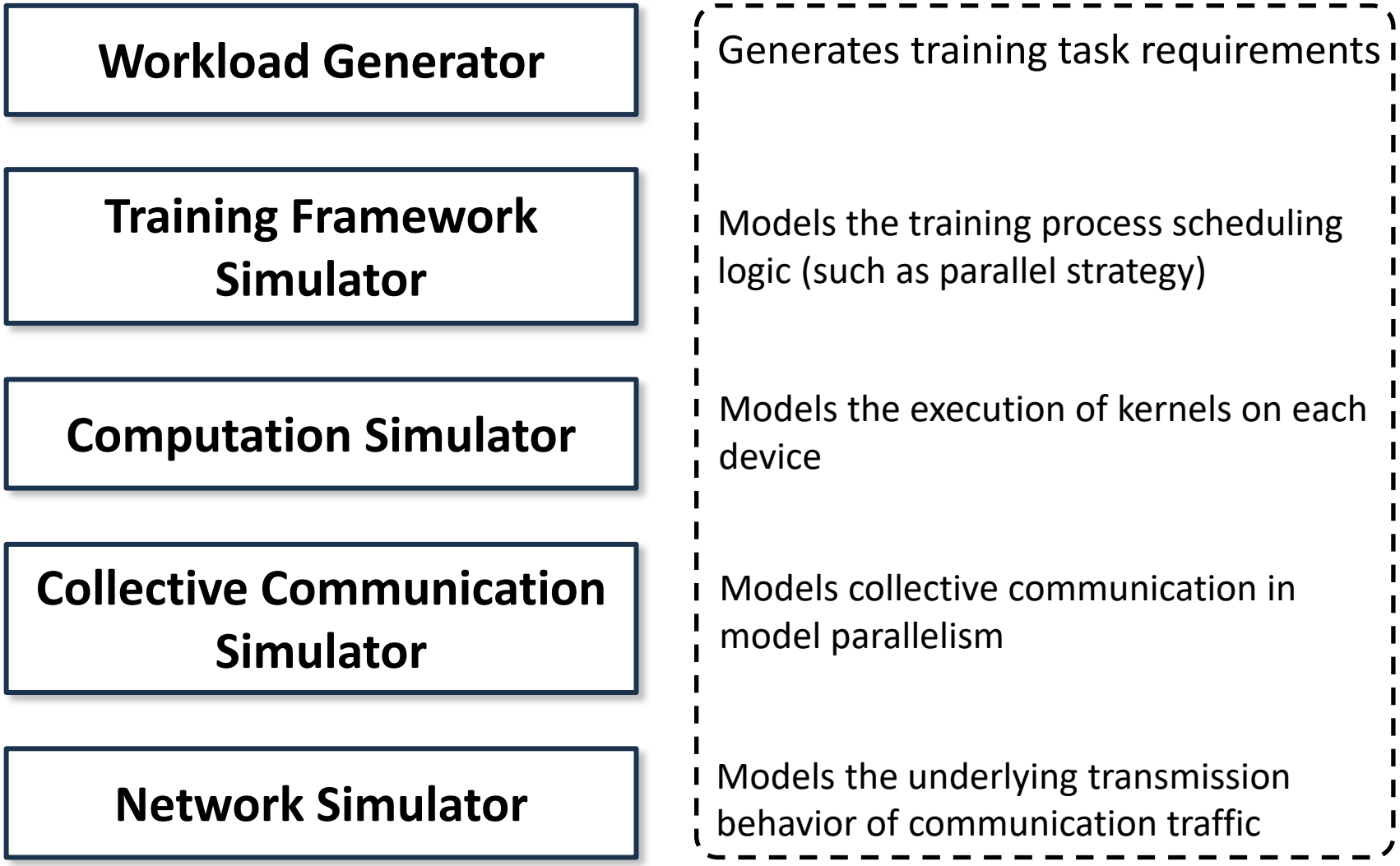
Design space of AI training is huge

Parallel Strategy	TP	SP	EP	PP	DP
Transport	DCQCN	HPCC	PFC	IRN	
Topology	Clos	Torus	Dragonfly	UB-Mesh	
…					

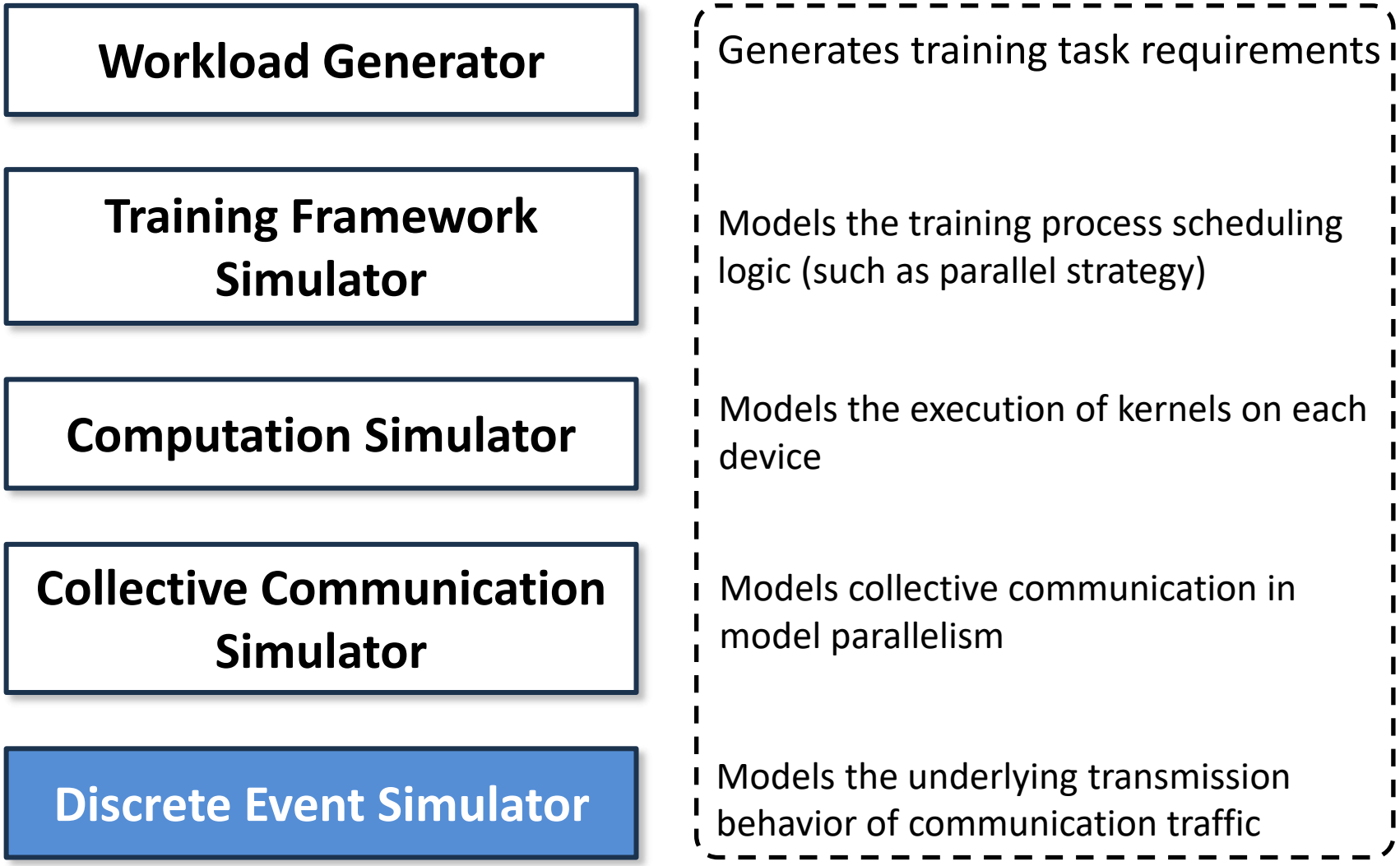
Simulation offers high flexibility and cost-effectiveness

- A *10,000* cluster costs more *than 1 million* dollars for per day

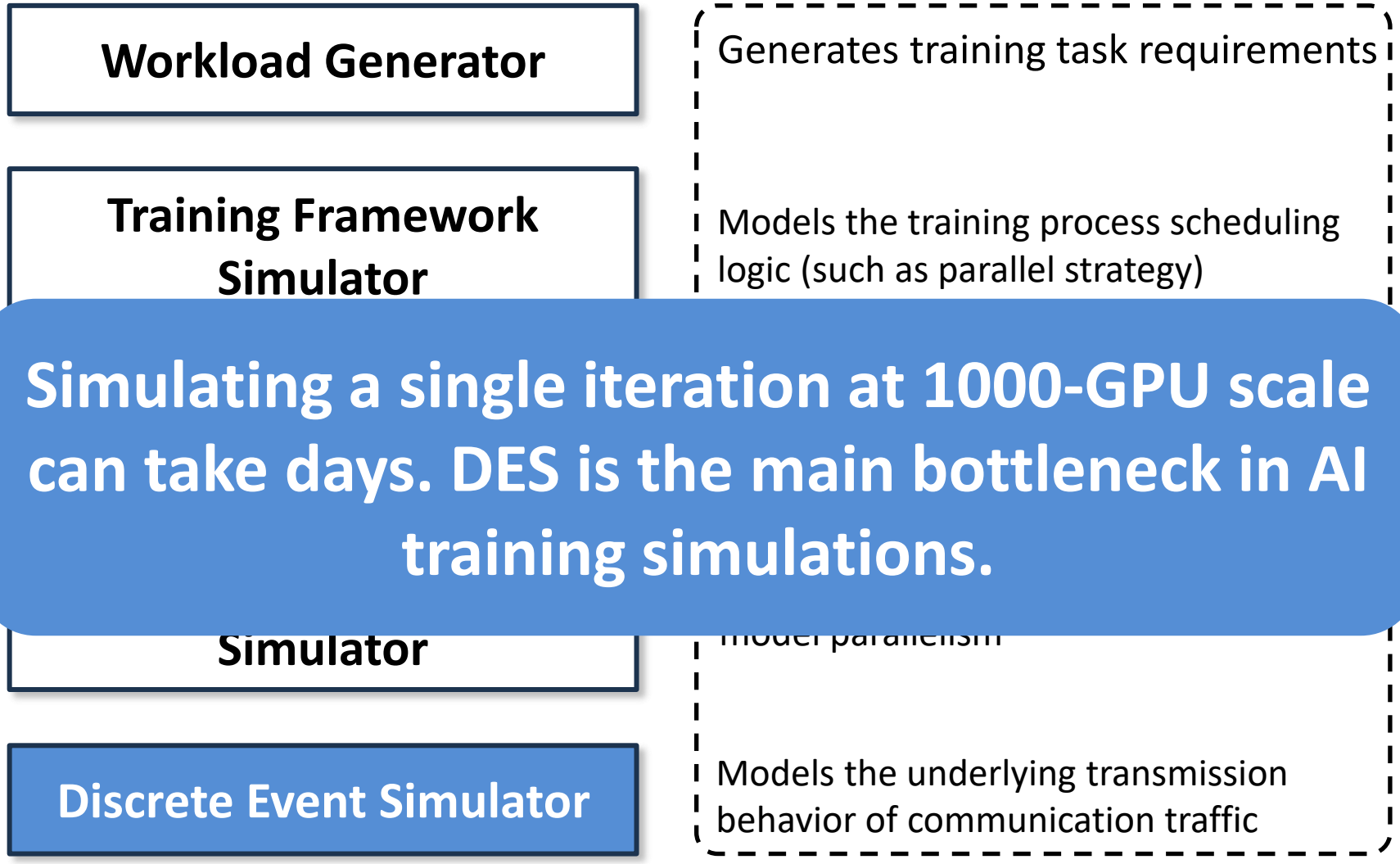
End-to-end AI Training Simulator Framework



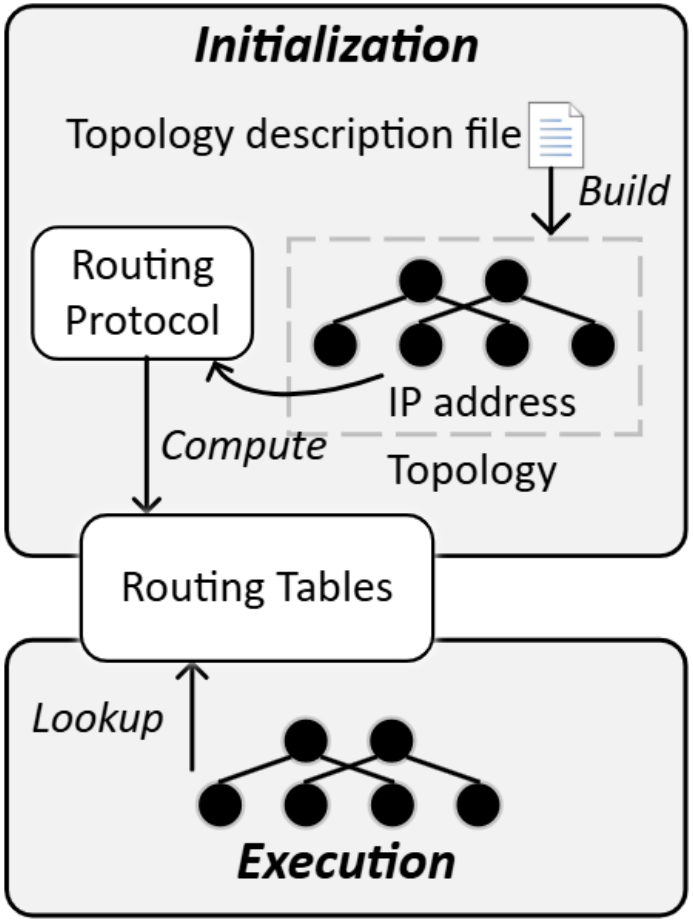
End-to-end AI Training Simulator Framework



End-to-end AI Training Simulator Framework



Network DES Workflow

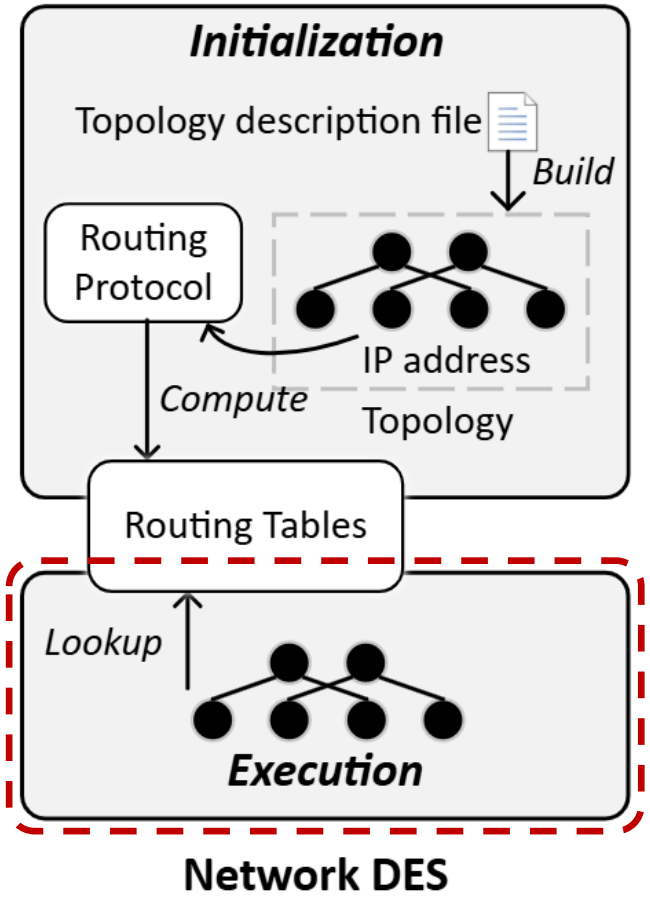


Network DES

- Build topology
- Compute routing tables

- Packet sending, queuing, forwarding and receiving

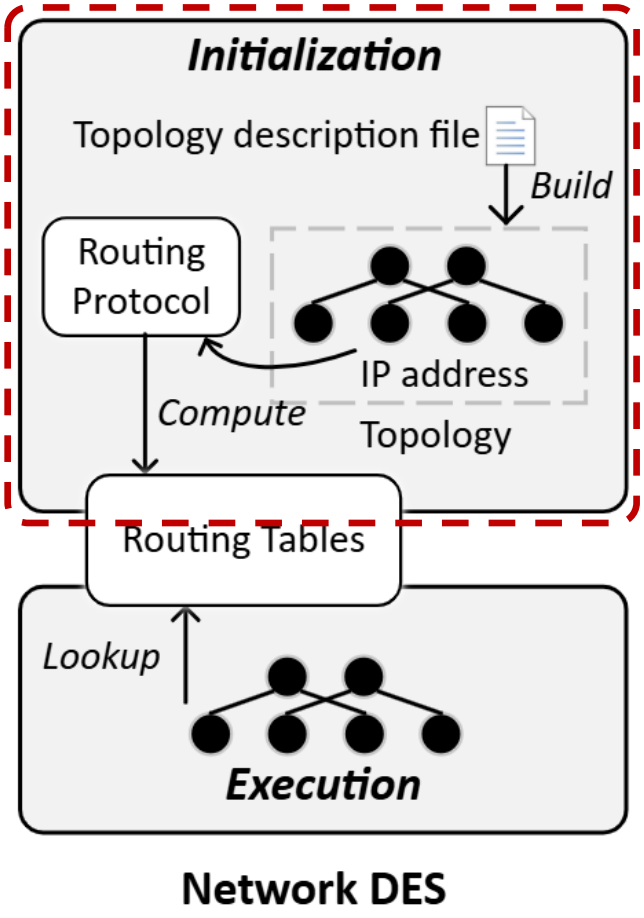
Bottleneck of Network DES



- Recent work try to improve parallelism in execution phase
  - *DONS*[1] data-oriented design
  - *UNISON*[2] fine-grained partitioning and load-adaptive scheduling
  - *SimAI*[3] lock-free sharing of global variables
  - *Multiverse*[4] sing process multiple experiments (SPME) and GPU acceleration

[1] Gao et al., SIGCOMM 2023
[2] Bai et al., EuroSys 2024
[3] Wang et al., NSDI 2025
[4] Gui et al., NSDI 2025

Bottleneck of Network DES

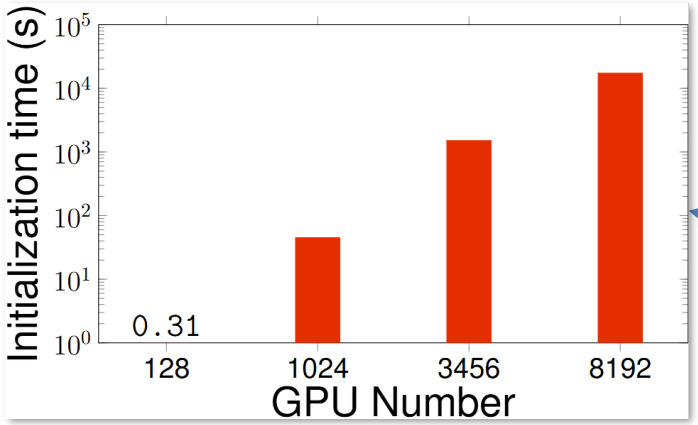


Network DES

- Our observation: control plane can affect simulation efficiency much
  - High computational overhead
  - High memory consumption
  - Inefficient routing table lookup
- These problems are orthogonal with previous work targeting improving parallelism

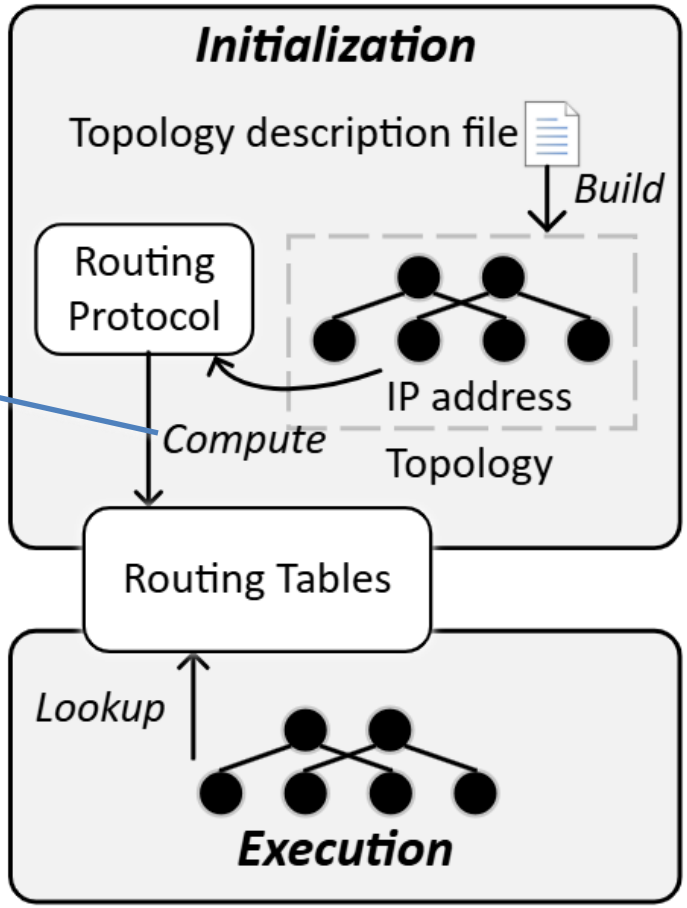
Bottleneck of Network DES

Long initialization time



It takes *5 hours* to compute routing tables for an 8K GPU Fattree in ns-3

The *ns-3-datacenter[1]* optimizes routing table computation but still struggle to support 10K+ scale cluster

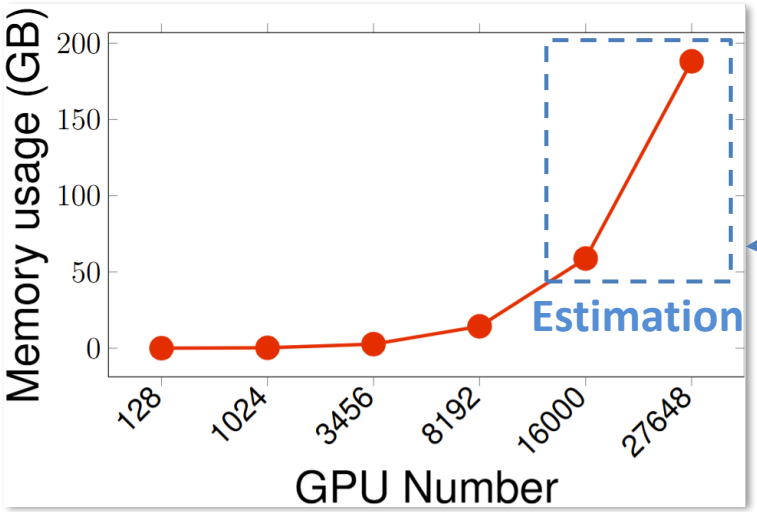


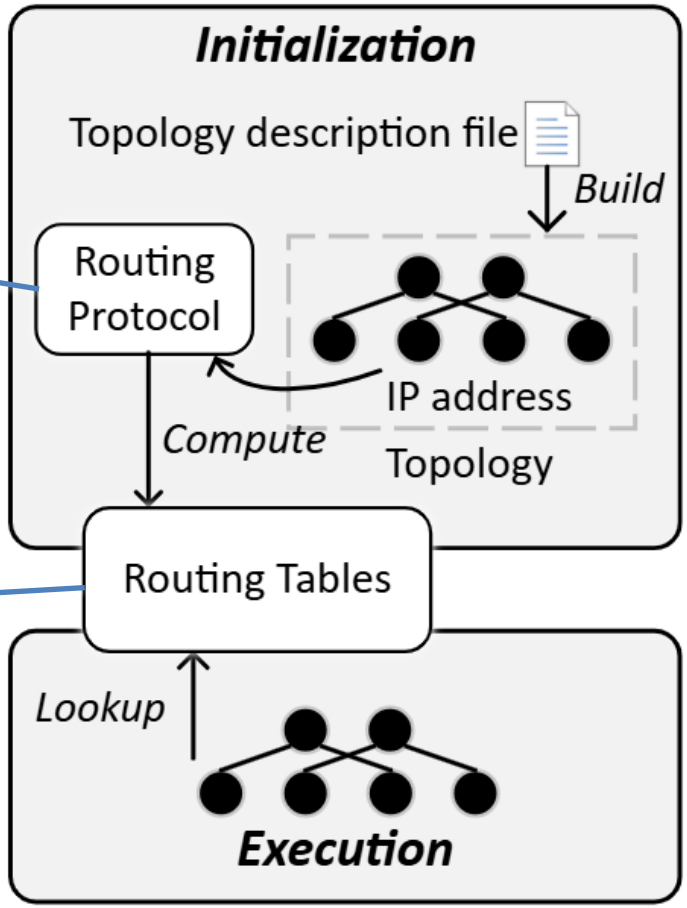
[1] Ns-3-datacenter. https://github.com/inet-tub/ns3-datacenter

Bottleneck of Network DES

Large memory consumption

Memory that used to maintain routing protocol states.





Network DES

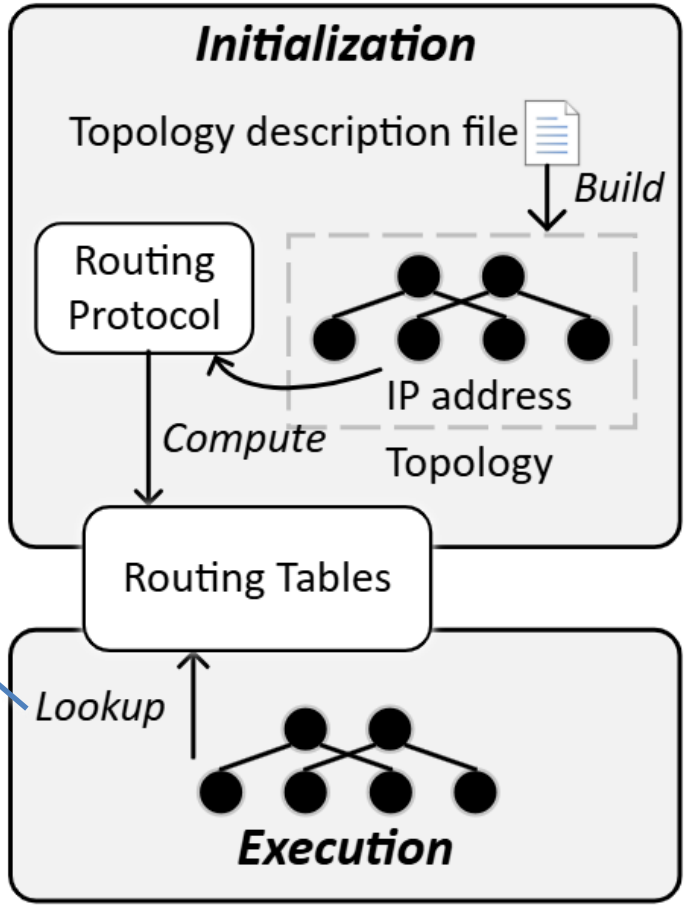
The ns-3 memory usage for routing tables is estimated to reach nearly 200 GB at a 27,000 GPU scale

Bottleneck of Network DES

Inefficient lookup

GPU Number	128	1024	3456
Lookup time(s)	21.06	1930.95	23256
Execution time(s)	42.90	2100.16	23808
Ratio (%)	49.09	91.94	97.68

Looking up account for over 97% of the total execution time in ns-3



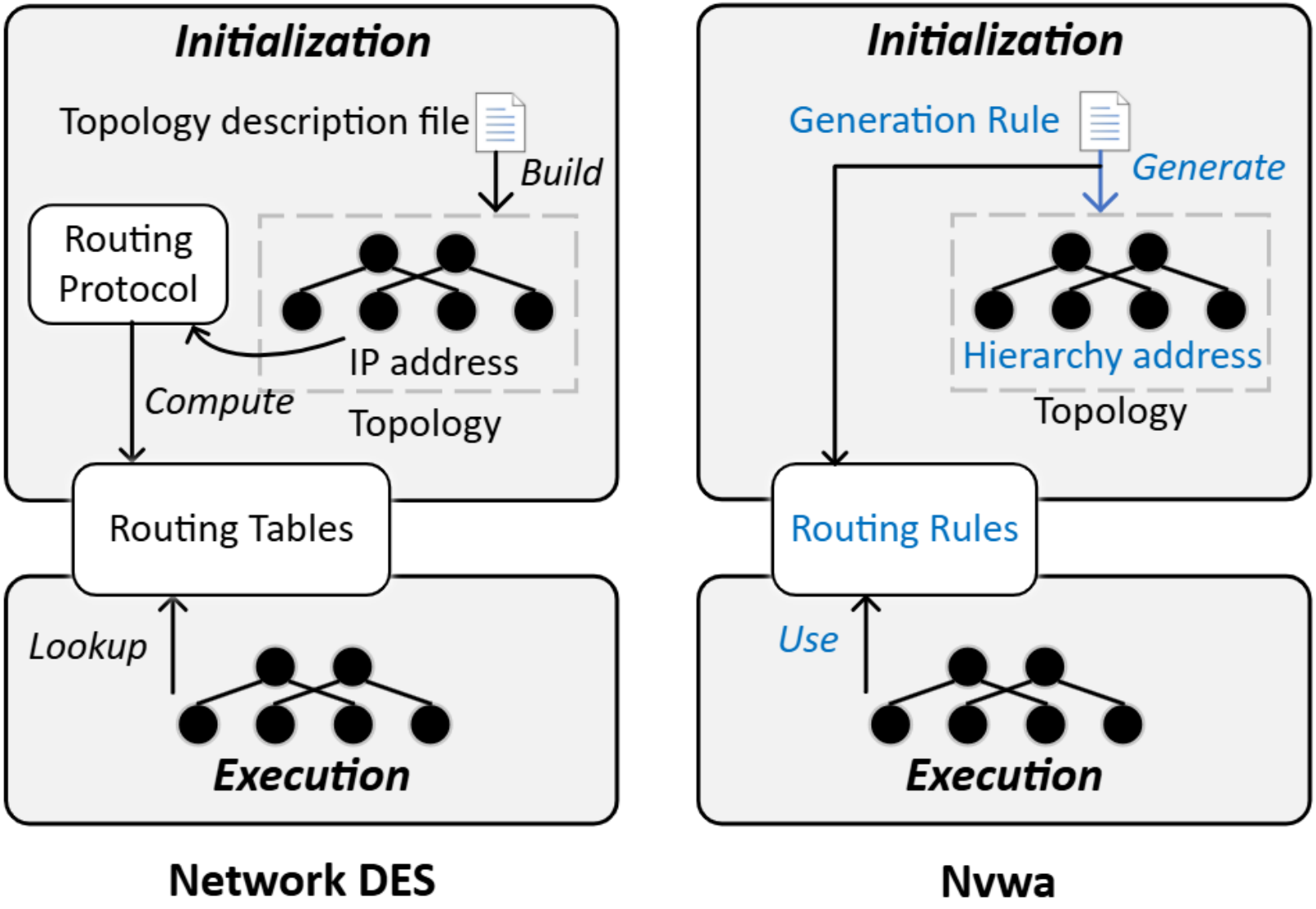
Network DES

Key Insight

- AI training clusters typically adopt regular and structured topologies
  - Clos, dragonfly, UB-Mesh
- The routing policies configured in these environments are relatively simple
  - Shortest path and near shortest path with simple policies
- As a result, routing paths exhibit highly regular patterns
  - E.g., up-and-down routing in Clos network

Idea: use routing rule to replace routing table for forwarding

Nüwa Architecture



Nüwa Workflow

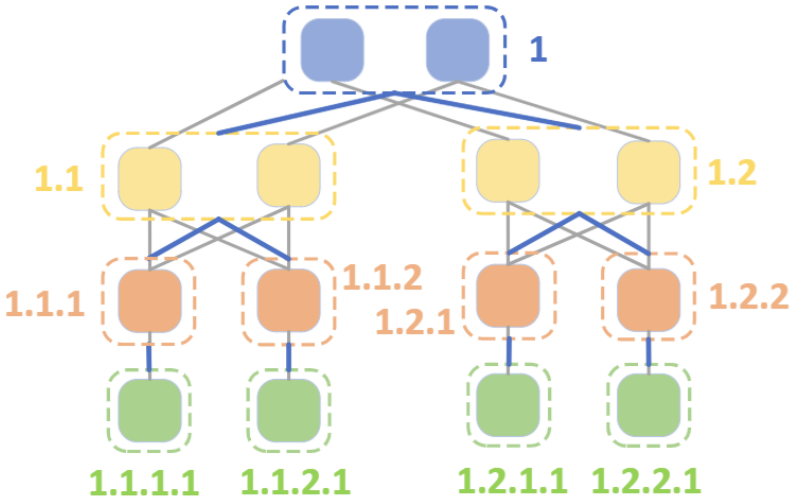
Core(2,1,2), SingleConnect

Agg(2,1,2), FullConnect

Edge(1,1,1), FullConnect

GPU(1,1,0)

Generation Rule

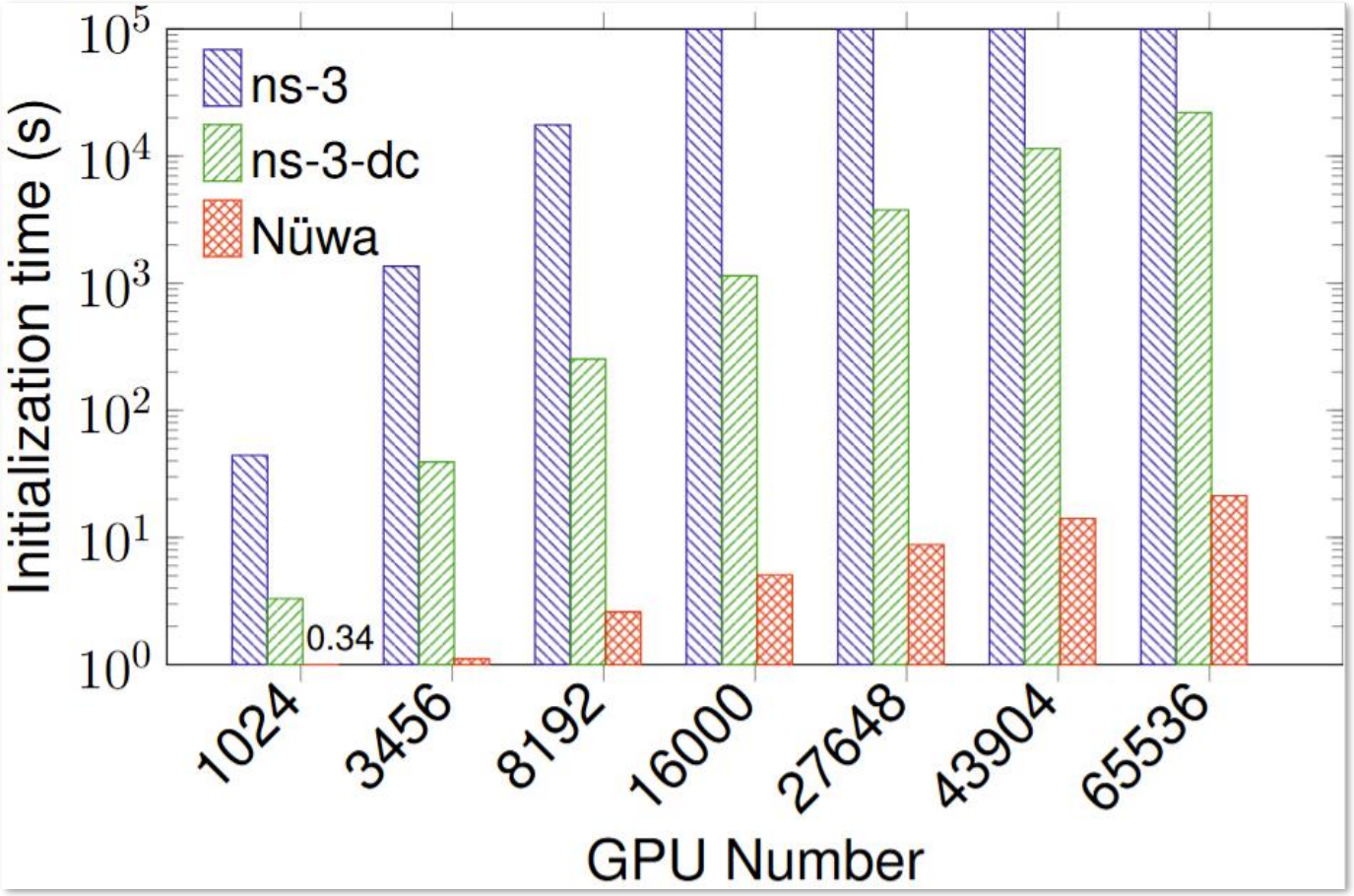


Hierarchy address

```
Up-and-down Rule (curA, dstA) :
    if curA is prefix of dstA :
        forward down to the subnet
    else :
        forward up
```

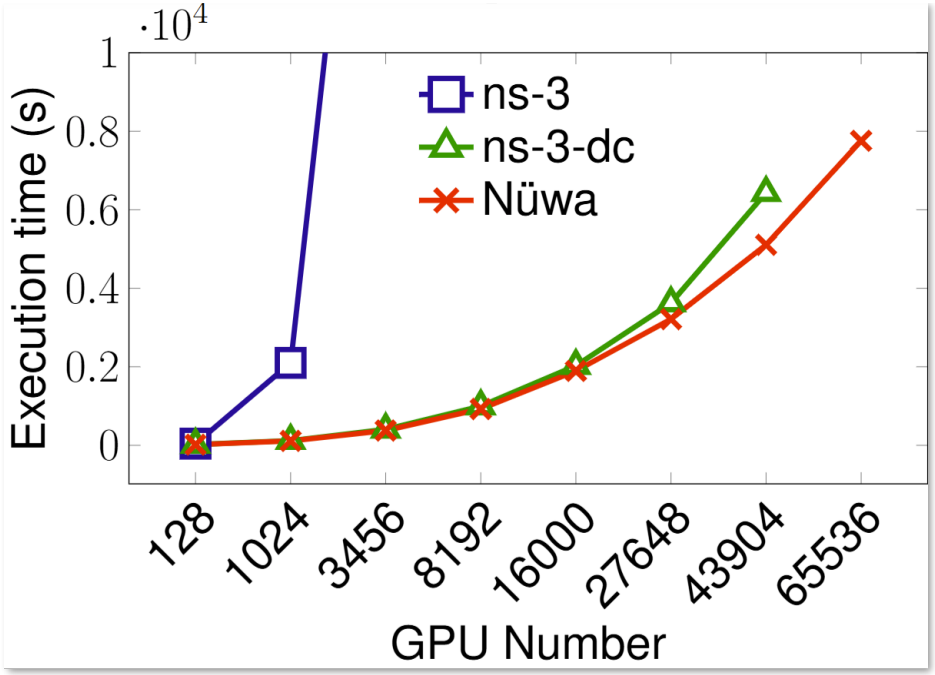
Routing Rule

Initialization Efficiency

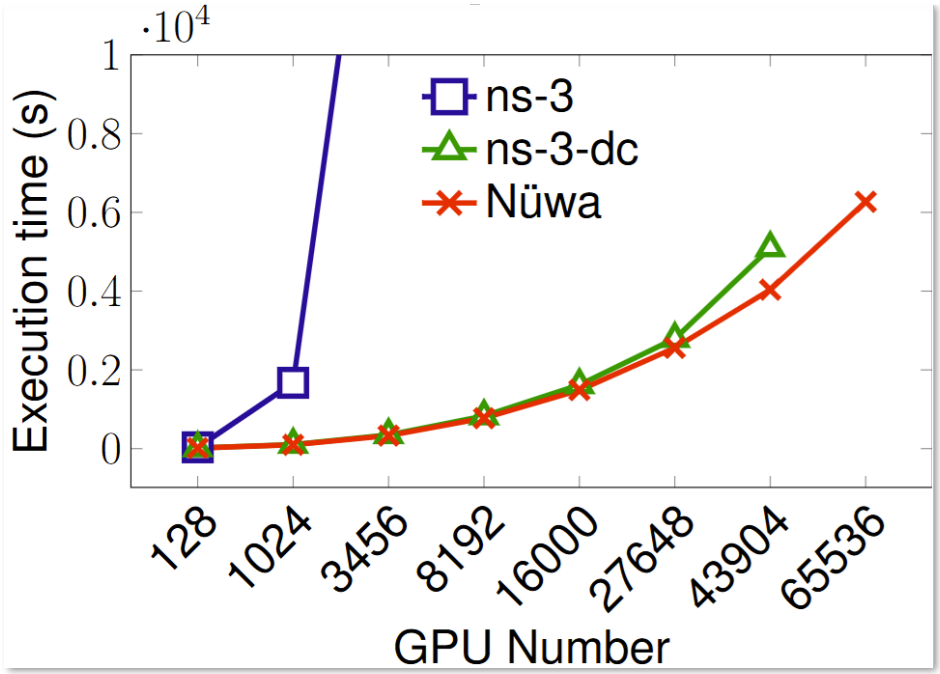


Nüwa takes *21s* to complete the initialization, which is *1000x* faster than ns-3-dc

Execution Efficiency



All-reduce



All-to-all

Nüwa's execution efficiency is higher than ns-3 and ns-3-dc, achieving up to 20% speed-up compared to ns-3-dc

Lookup Efficiency

Lookup time in all-reduce execution

GPU Number	1024	3456	8192	16000	27648	43904	65536
ns-3 (s)	1930.95	23256	1 day[1]	1 day	1 day	1 day	1 day
ns-3-dc (s)	2.01	9.11	22.88	59.36	100.87	269.13	OOM[2]
Nüwa (s)	0.012	0.02	0.12	0.19	0.43	0.93	1.12

[1]Denotes simulation time is more than 1 day.
[2]Denotes simulation out of memory.

Nüwa's lookup is 200x faster than ns-3-dc, almost negligible

Future Work

- Support other topologies
- Support other simulators
- Simulate failures in network
- We will make Nüwa open source

Dragonfly

UB-Mesh

OMNet++

DONS

Conclusion

- Initialization of the control plane becomes a bottleneck in DES network simulator
- Nüwa achieves efficiency and scalability by replacing the routing table with high-level routing rule
  - Achieves negligible initialization time
  - Saves a lot of memory consumption
  - Speeds up execution

Nüwa: Efficient Generative Control Plane for AI Network Simulation

Thanks!

wkli24@stu.xjtu.edu.cn