GRU and EdgeQ-Learning based Traffic Prediction and Scaling of SFC

Suman Pandey, Jae-Hyoung Yoo, James Won-Ki Hong

ACM SIGCOMM N2Women 2021, August 23

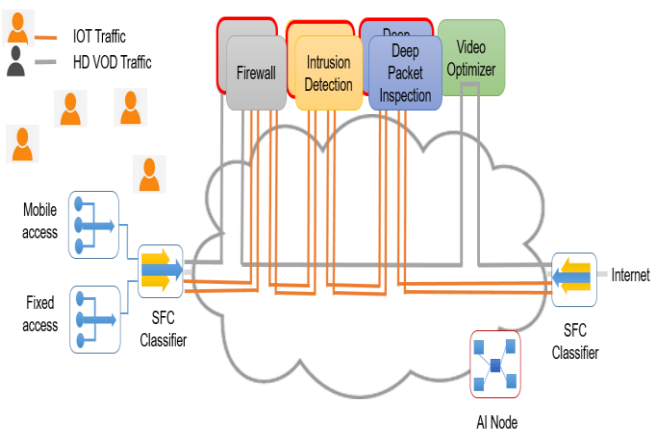
Goal

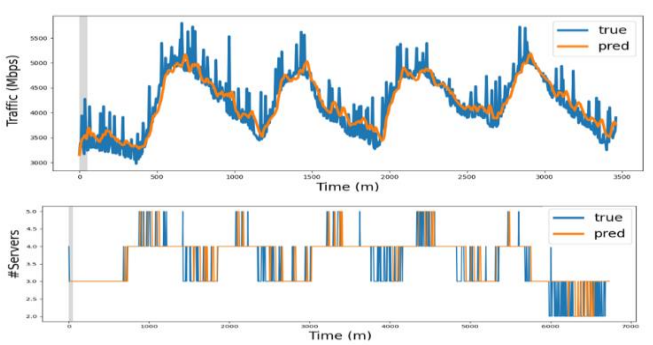
Provide AI based Optimal Scaling and Placement

- **Increase the Elasticity of the Network Services**
- **While ensuring SLA in terms of Throughput, End-to-End Delay, Successful serving of request**
- **Reduce Scaling cost, by reducing number of scaling operation**

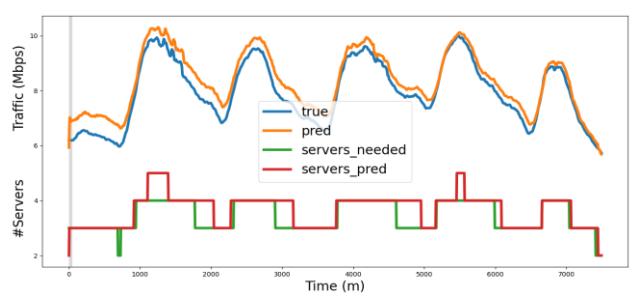
Scaling (GRU based Predictive Scaling)

Monitor Resource utilization of VNF
CPU , Memory , Disk , I/O , # of requests
Proposed Approach – GRU based Proactive Scaling





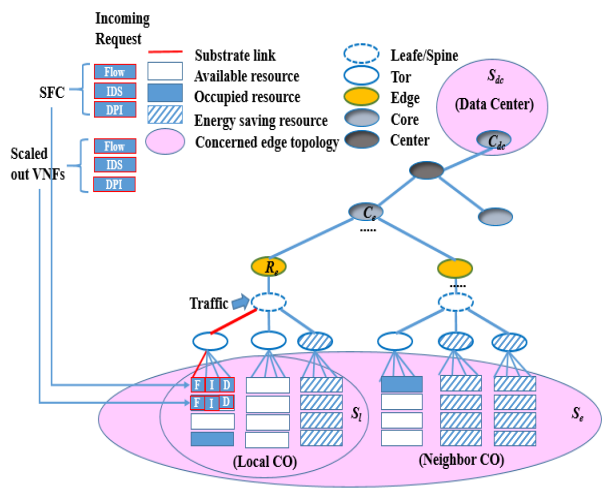
Test bed Traffic and Scaling



Simulated Traffic and Scaling

Placement (Edge-QLearning)

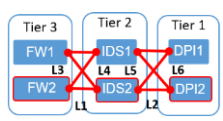
Monitor underlying Infrastructure
CPU , Memory, Bandwidth, Location of servers
Proposed Approach - Q-Learning based placement



States (s) – VNFs in SFC
Actions (a) – Servers for placement
State transition probability (pi) – Resource availability on server

Rewards (r) - $r = e^{-\frac{\beta l}{z}}$

l = L1 + L2 + L3 + L4 + L5
z = length of the SFC



Q-Table

Prioritize edge

Extend to Neighbor & DC

Servers (actions)

SFC (States)

Q	Local		Neigh		DC	
vnf 1	0.0545	....	0.1051	....	0.0545	....
vnf2	0.0655	.....	0.0245	.....	0.0655	.....
:	:	:	:	:	:	:
vnf N	0.0n	0.0n	0.0n	0.0n	0.0n	0.0n

If not successful — Second hierarchy

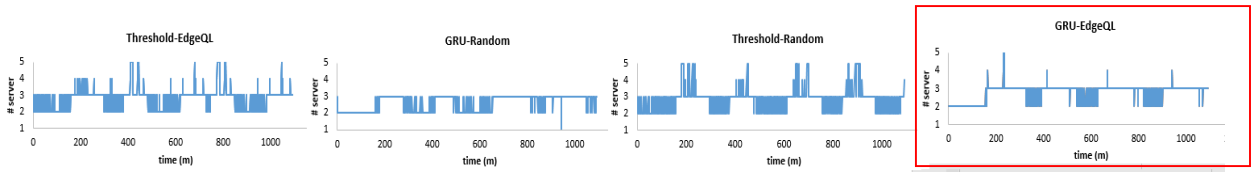
Q	Local		Neigh		DC	
vnf 1	0.0545	....	0.1051	....	0.0545	....
vnf2	0.0655	.....	0.0245	.....	0.0655	.....
:	:	:	:	:	:	:
vnf N	0.0n	0.0n	0.0n	0.0n	0.0n	0.0n

If not successful — Third hierarchy

Q	Local		Neigh		DC	
vnf 1	0.0545	....	0.1051	....	0.0545	....
vnf2	0.0655	.....	0.0245	.....	0.0655	.....
:	:	:	:	:	:	:
vnf N	0.0n	0.0n	0.0n	0.0n	0.0n	0.0n

Results

	Threshold-random	GRU-random	Threshold-EdgeQL	GRU-EdgeQL
Avg latency (ms)	159.76	98.70	106.99	**98.09**
SLA violation %	11.87	10.37	10.05	**9.42**
Scaling Operation %	36.86	14.30	34.52	**10.5**



Related work
Suman Pandey at el, "EdgeDQN: Multiple SFC Placement in Edge Computing Environment", 17th International Conference on Network and Service Management (CNSM 2021)
Suman Pandey at el, "GRU and EdgeQ-Learning based Traffic Prediction and Scaling of SFC", The 7th IEEE International Conference on Network Softwarization (NetSoft 2021)
Suman Pandey at el, "Environment Aware Adaptive Q-Learning to Deploy SFC on Edge Computing", 16th International Conference on Network and Service Management (CNSM 2020)
Suman Pandey at el, "Q-Learning Based SFC Deployment on Edge Computing Environment", The 21st Asia-Pacific Network Operations and Management Symposium (APNOMS 2020)