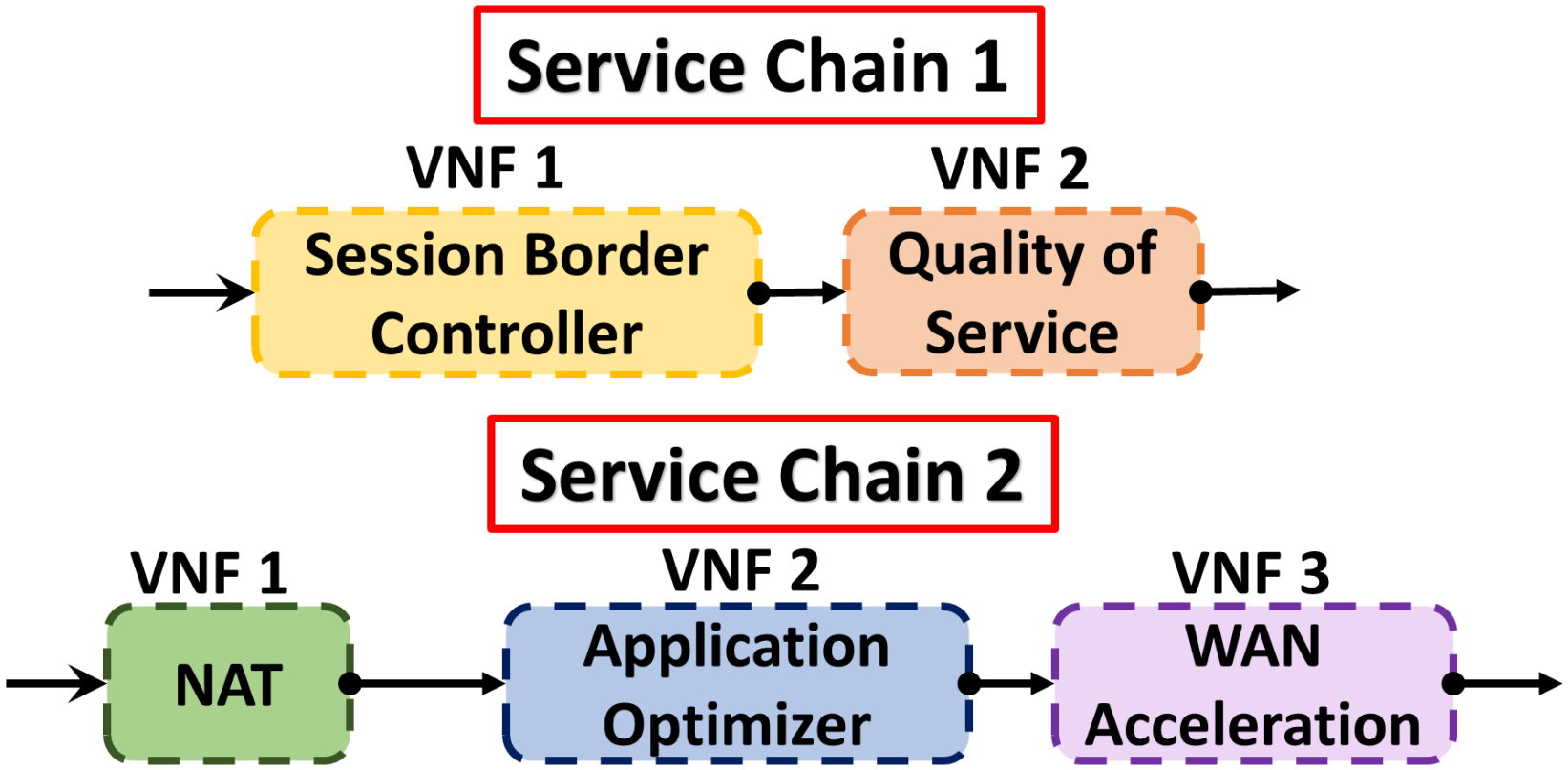
Deploying Multiple Service Chain (SC) Instances per Service Chain

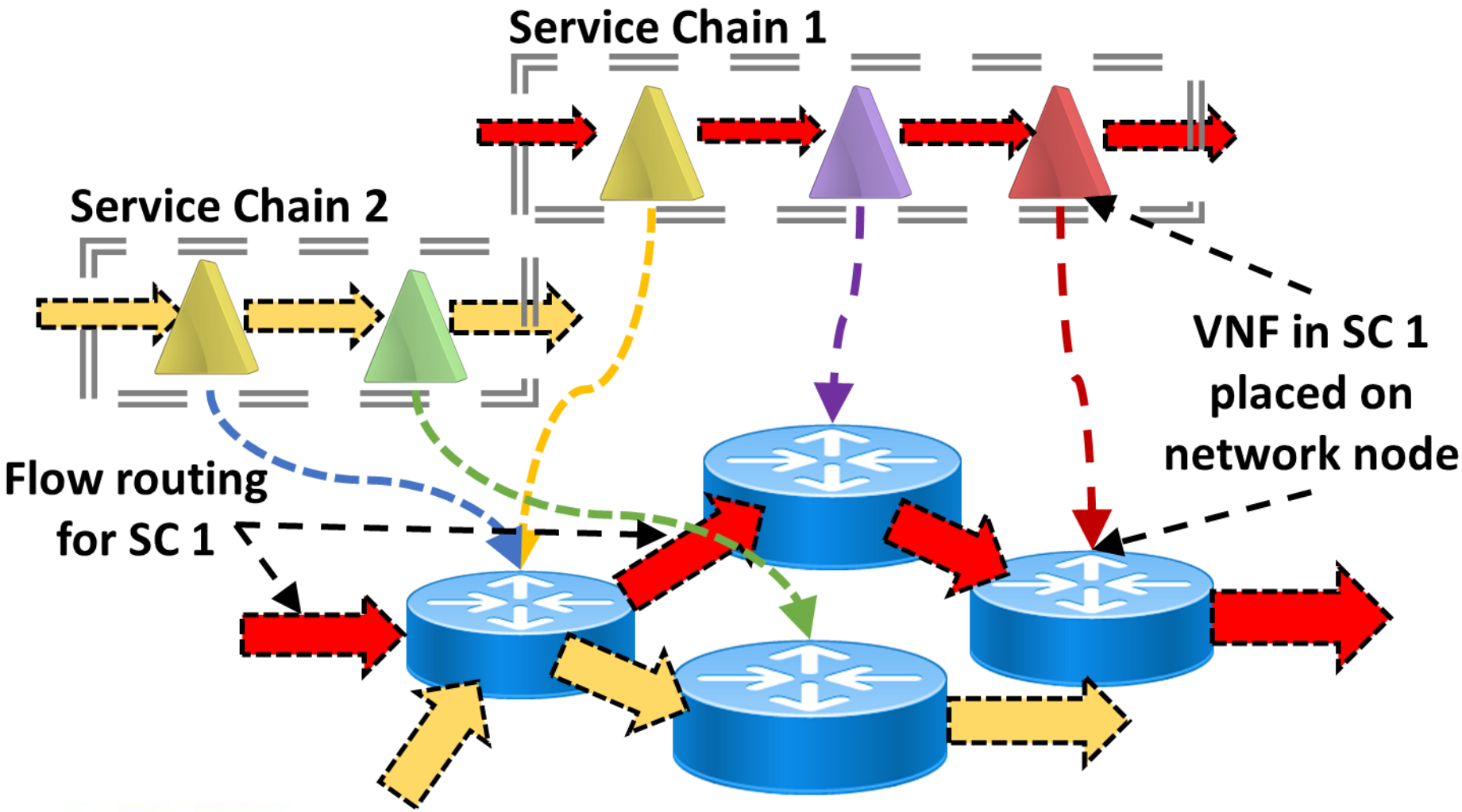
BY
ABHISHEK GUPTA
FRIDAY GROUP MEETING
JULY 14, 2017

UCDAVIS

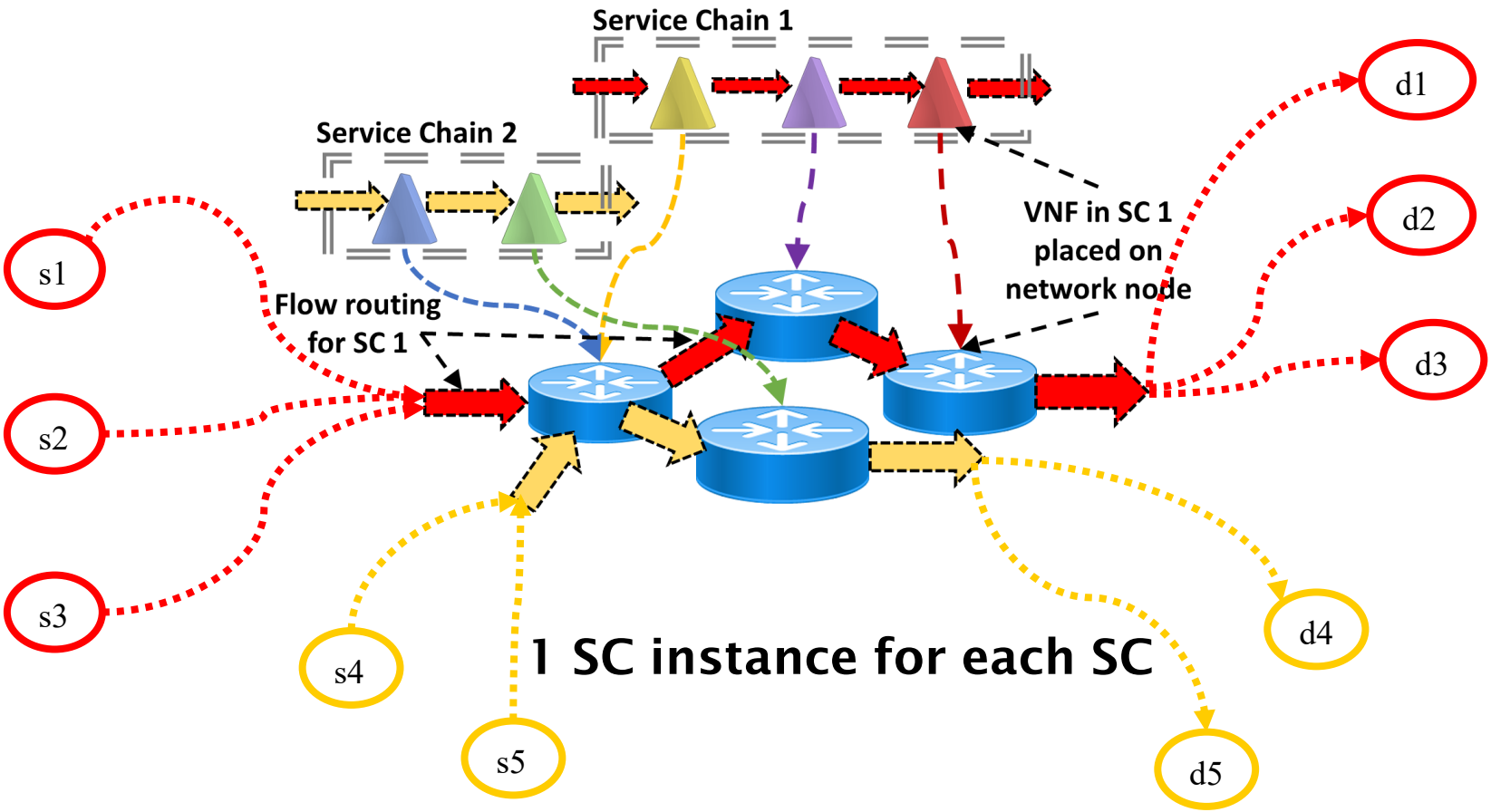
Virtual Network Function (VNF) Service Chain (SC)



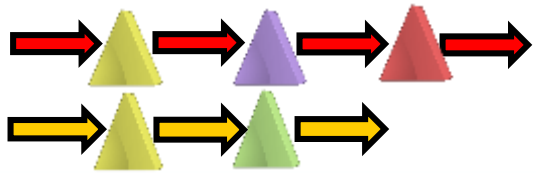
Multiple VNF SC Placement and Routing



Continued…


Service Chain 1
Service Chain 2
VNF in SC 1
placed on
network node
Flow routing
for SC 1
s1
s2
s3
s4
s5
d1
d2
d3
d4
d5
1 SC instance for each SC

Single Instance Per SC



Service Chain 1

Service Chain 2

2 Demand Requests SC1

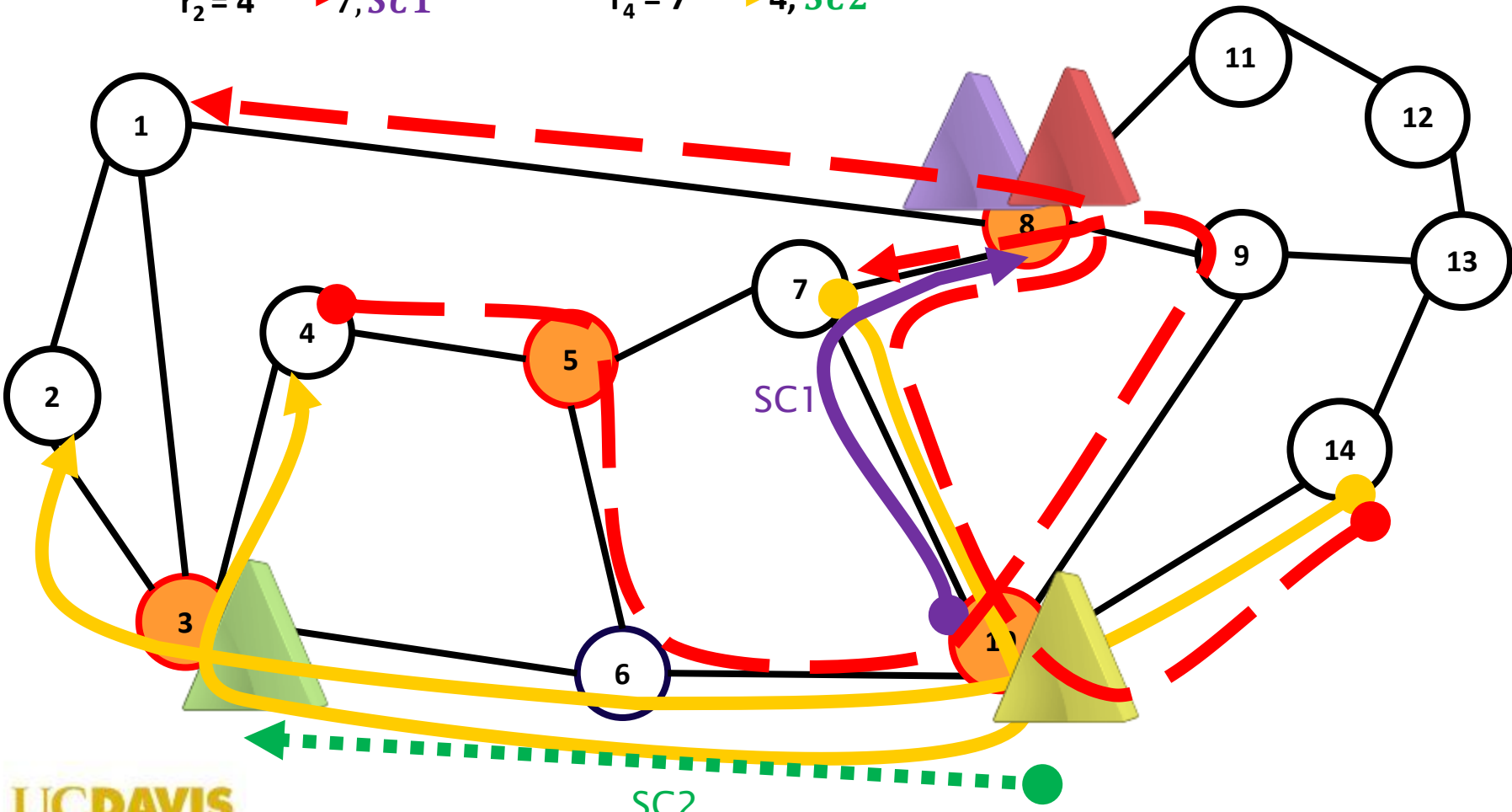
r_1 = 14 → 1, $SC1$

r_2 = 4 → 7, $SC1$

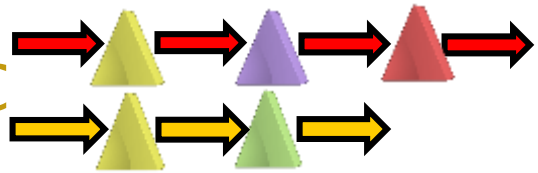
2 Demand Requests SC2

r_3 = 14 → 2, $SC2$

r_4 = 7 → 4, $SC2$



Multiple Instances per SC



Service Chain 1

Service Chain 2

2 Demand Requests SC1

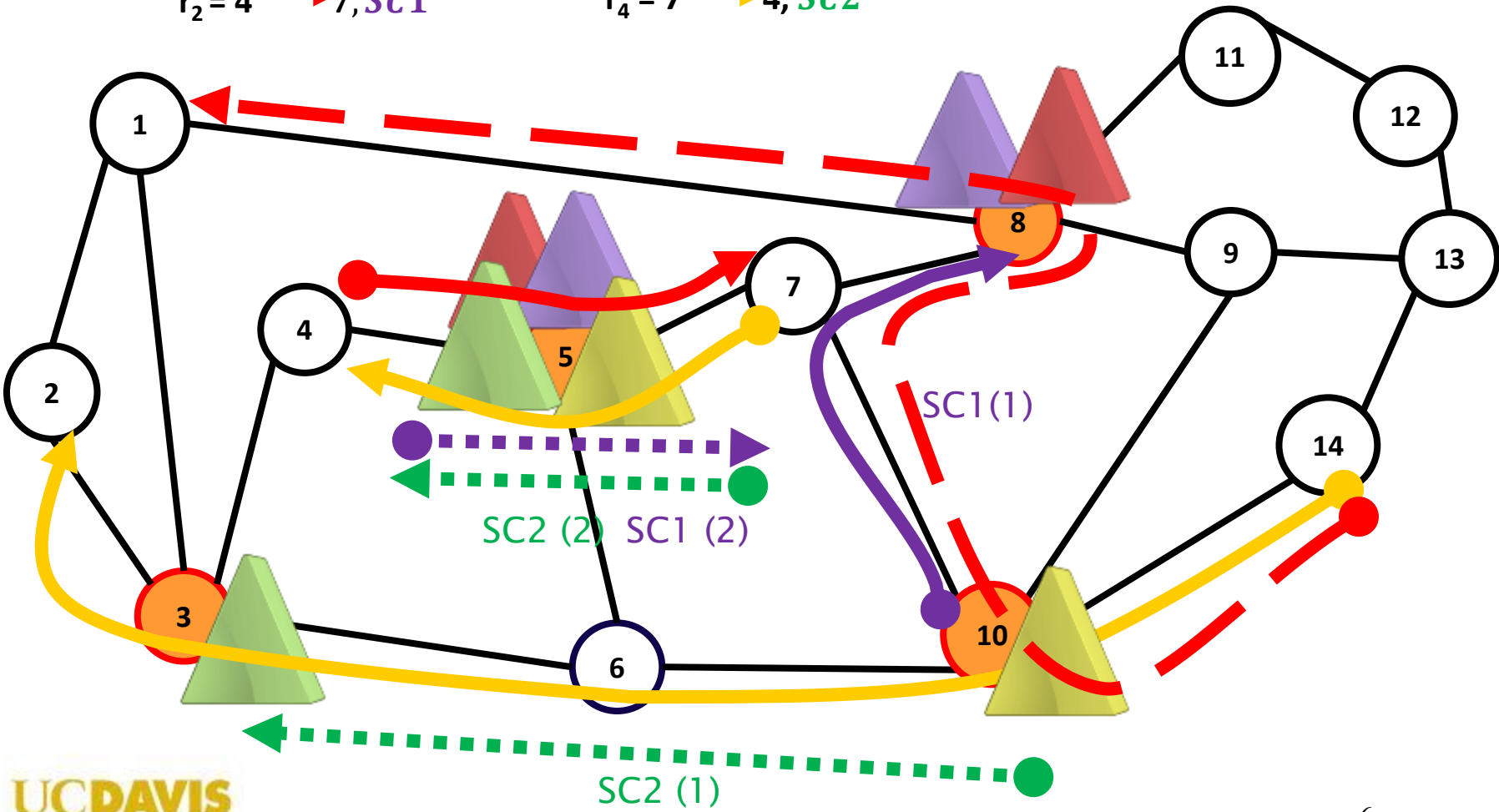
r_1 = 14 → 1, $SC1$

r_2 = 4 → 7, $SC1$

2 Demand Requests SC2

r_3 = 14 → 2, $SC2$

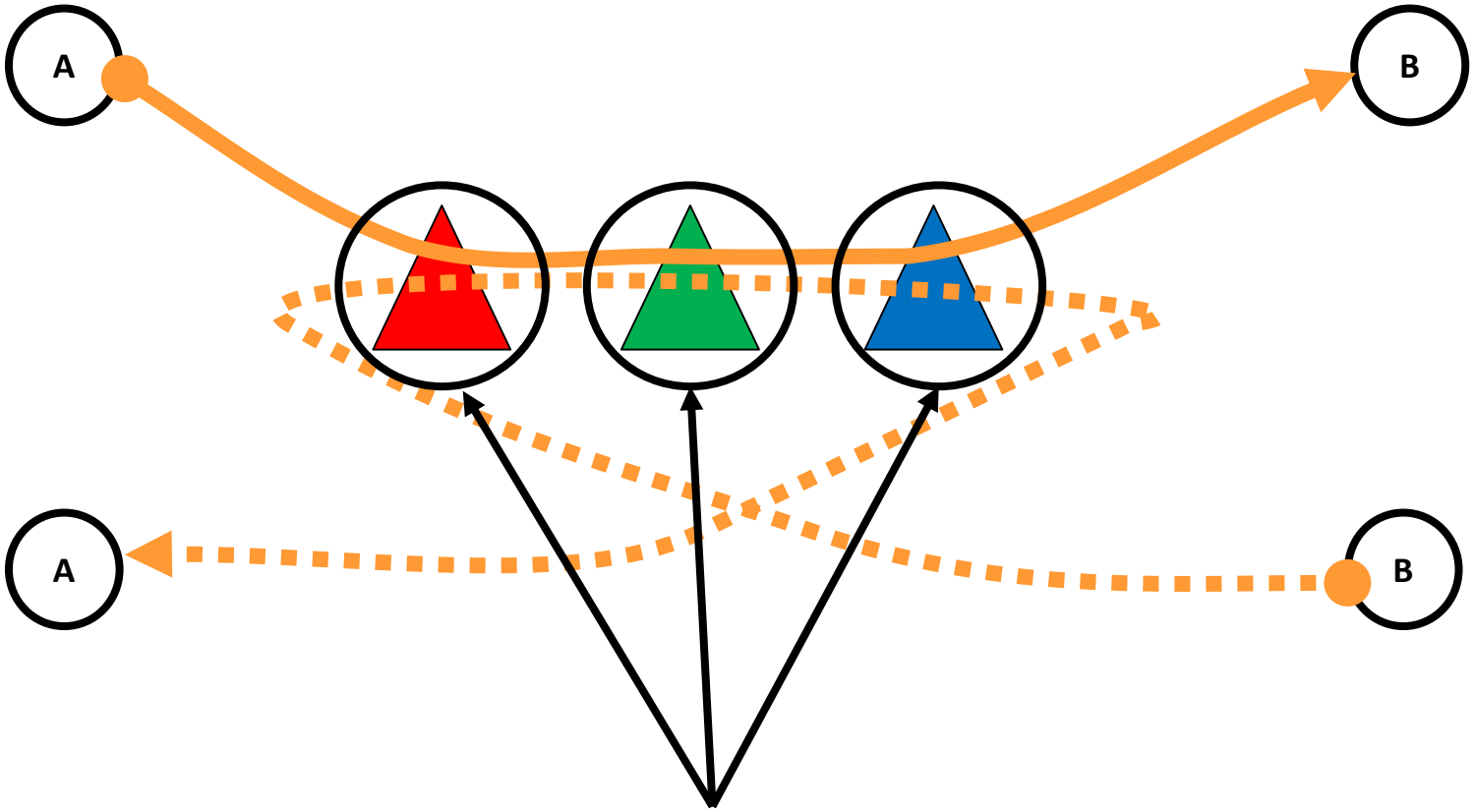
r_4 = 7 → 4, $SC2$



Inferences and Questions

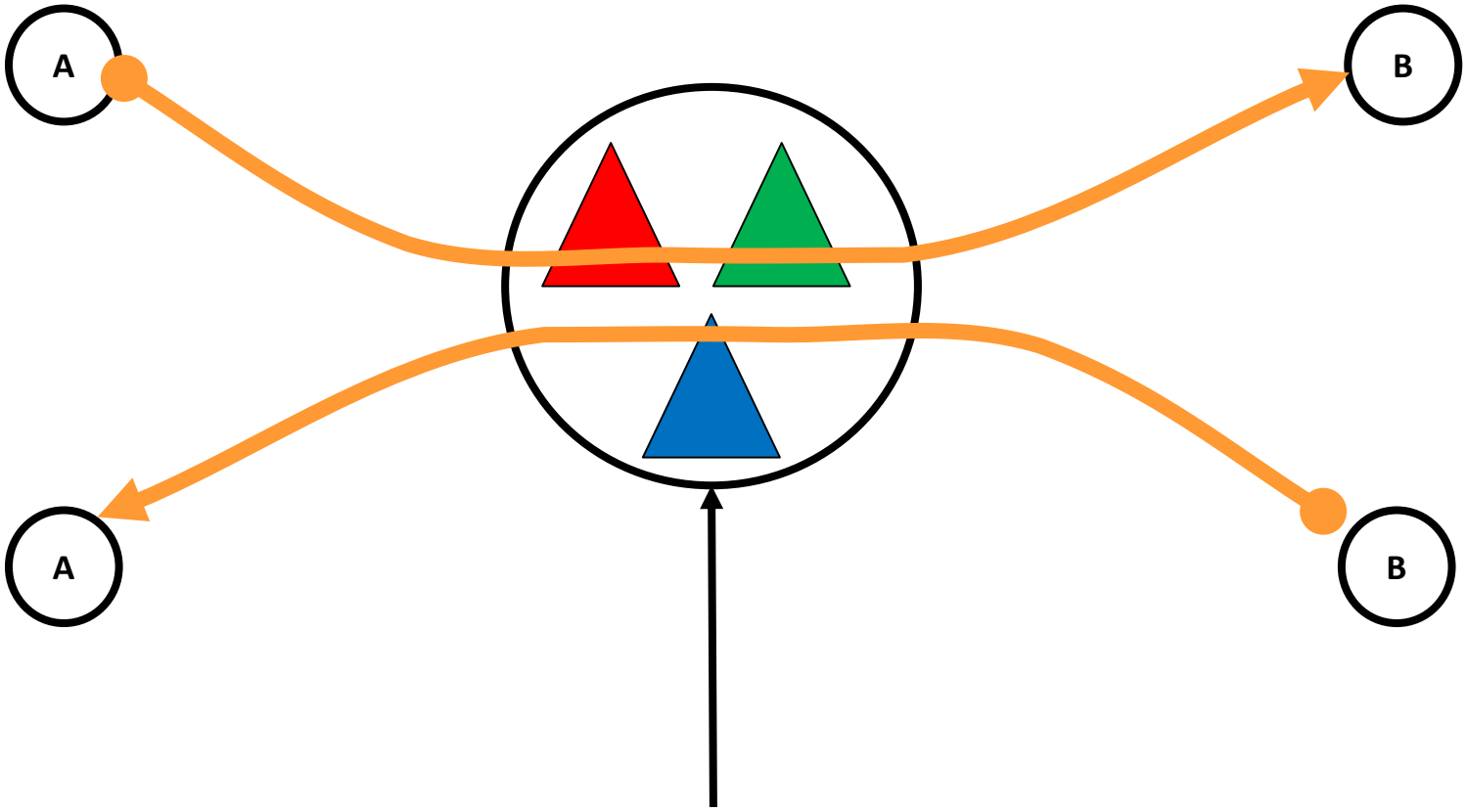
- 1 SC instance per SC leads to suboptimal results
- Having SC instances replicated on every node will lead to to optimal results
  - Large capital expenditure to make all nodes NFV capable
  - High Orchestration Overhead for large number of instances
- The question therefore becomes:
  - **How many SC instances to deploy to reduce bandwidth consumption while also reducing nodes used?**

Issue of symmetric flows



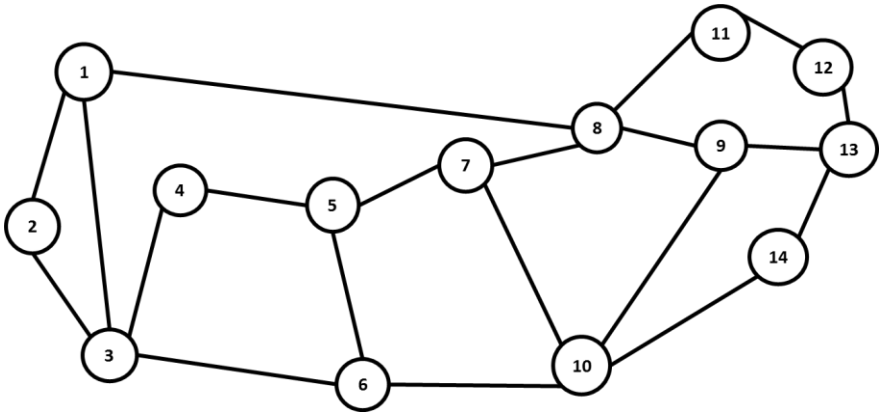
VNFs placed in different nodes

Continued…

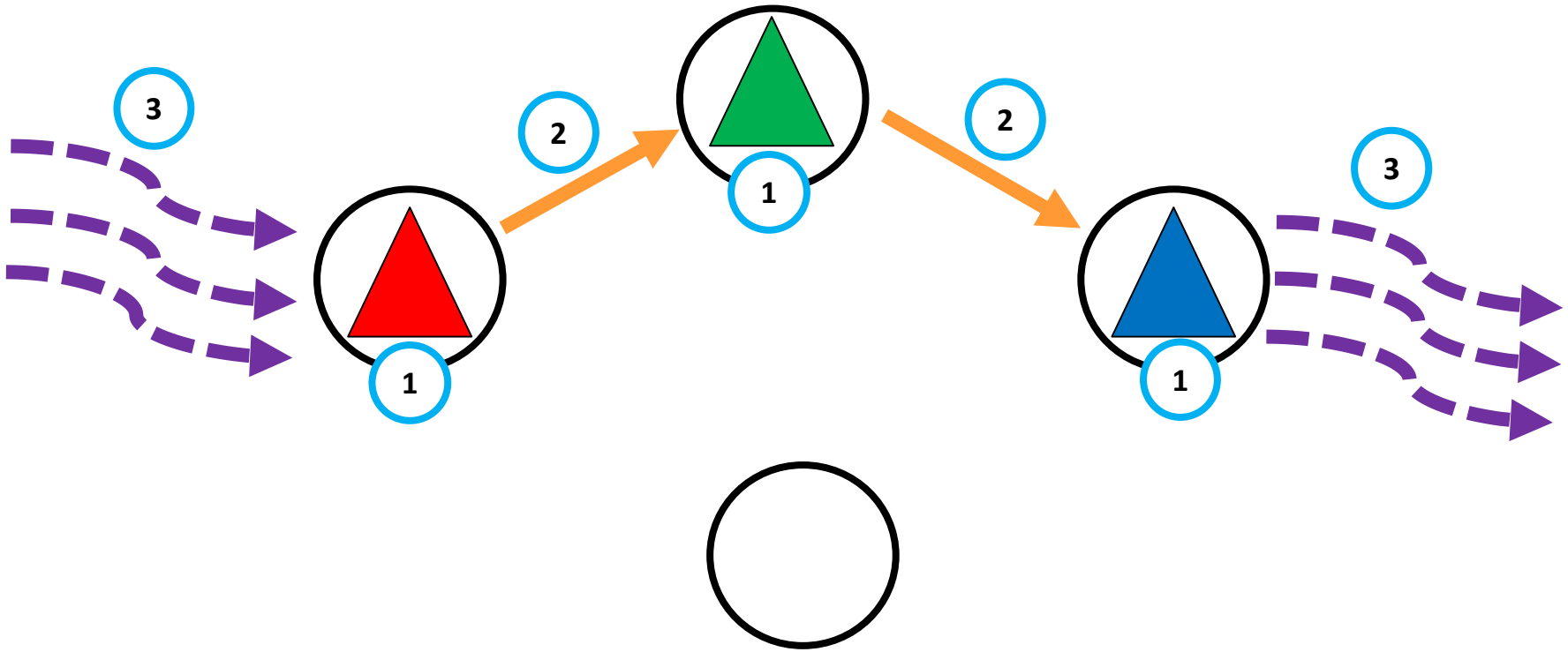


Continued…

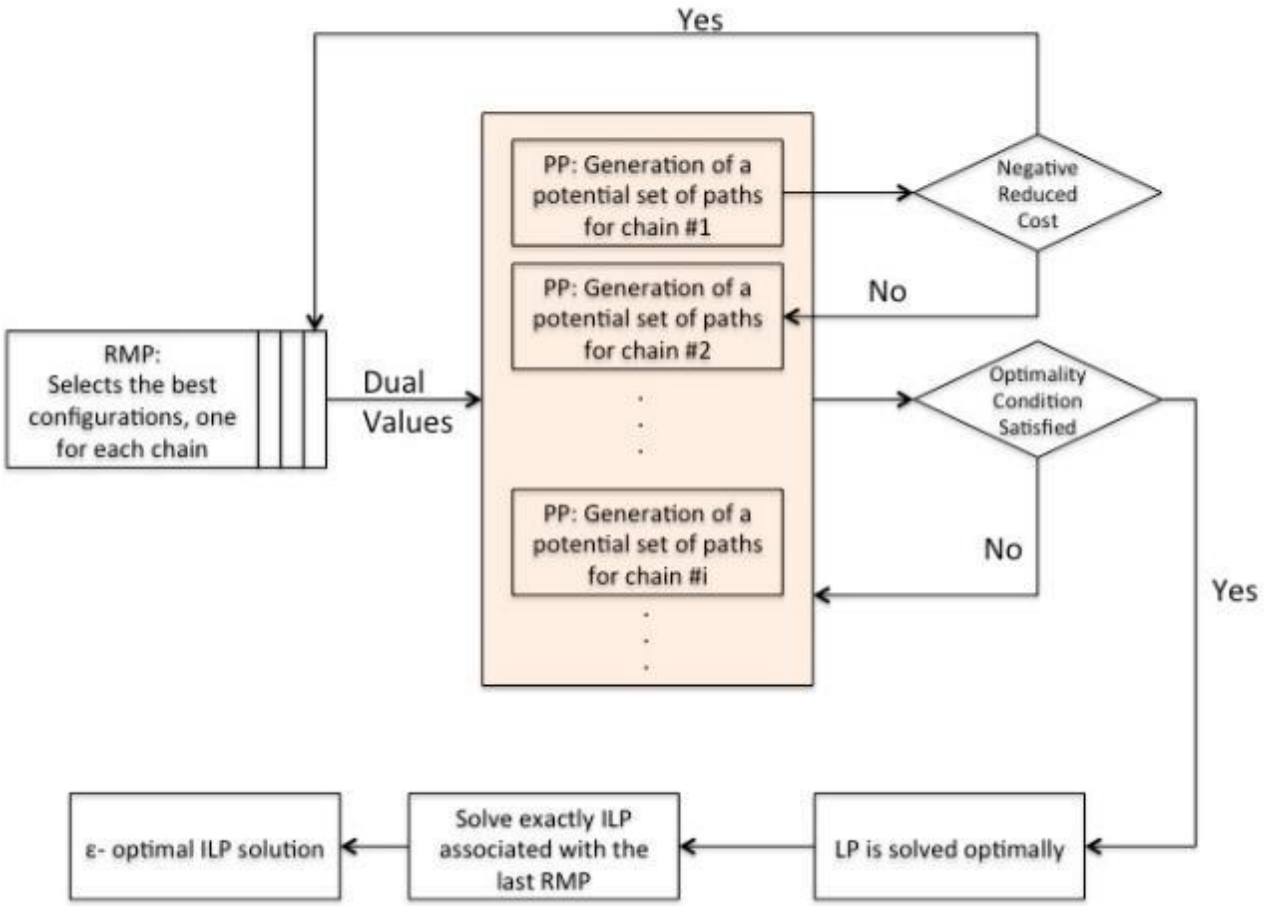
- Placing VNFs for SC at different nodes
  - makes symmetric flow take longer path
- Placing VNFs for SC at one node
  - symmetric flow takes shorter path
  - placement and routing becomes easier
  - chaining aspect is forgone
  - Is this more realistic?
  - Represents the case of a DC



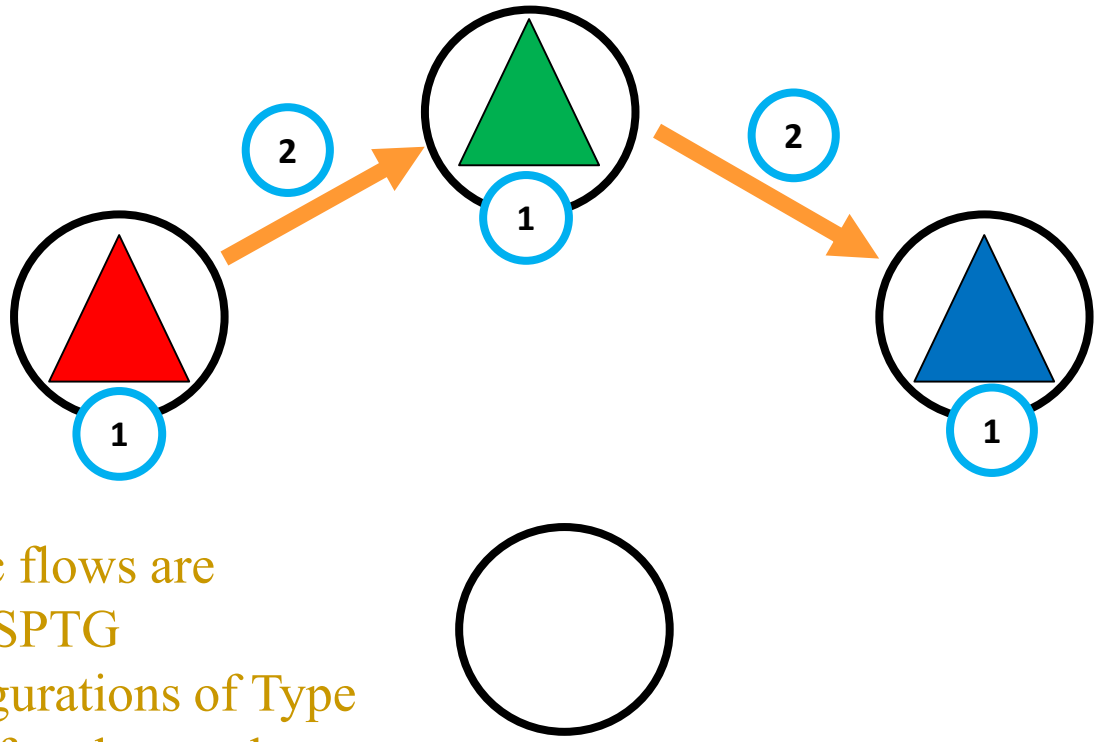
Configuration Type 1 – (ILP, CG)



CG Model


Yes
PP: Generation of a potential set of paths for chain #1
Negative Reduced Cost
PP: Generation of a potential set of paths for chain #2
No
RMP: Selects the best configurations, one for each chain
Dual Values
Optimality Condition Satisfied
PP: Generation of a potential set of paths for chain #i
No
Yes
ε- optimal ILP solution
Solve exactly ILP associated with the last RMP
LP is solved optimally

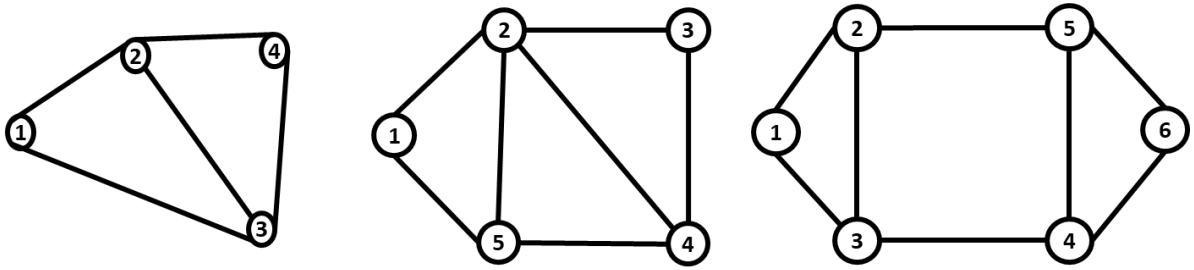
Configuration Type 2 – (2 Phase Model)

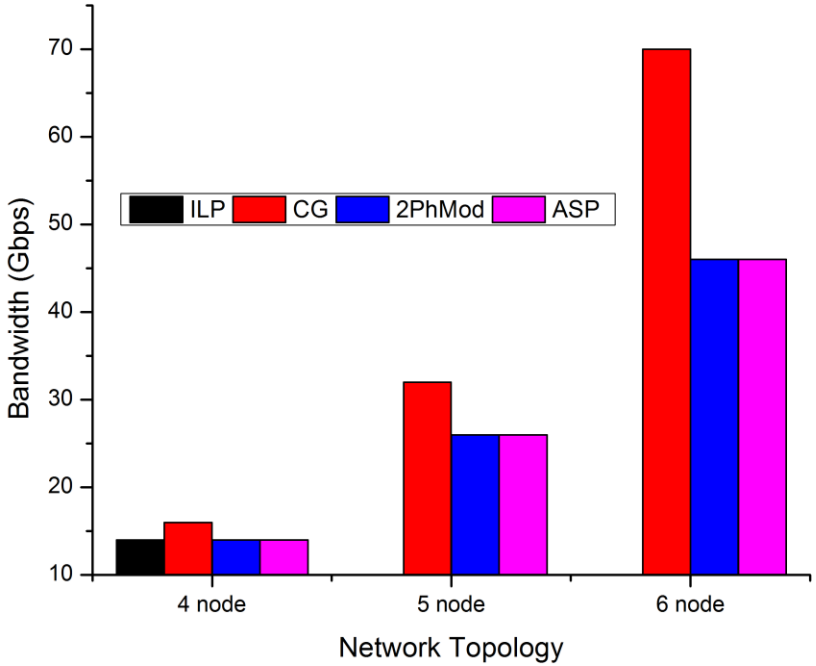


Phase 1 : Traffic flows are clustered using SPTG

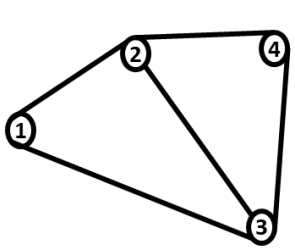
Phase 2 : Configurations of Type 2 are generated for clustered traffic of Phase 1

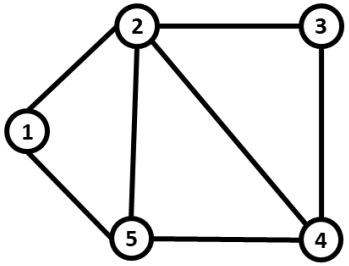
Comparison (ILP, CG, 2 Phase Model)

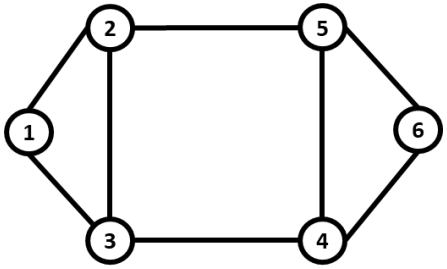


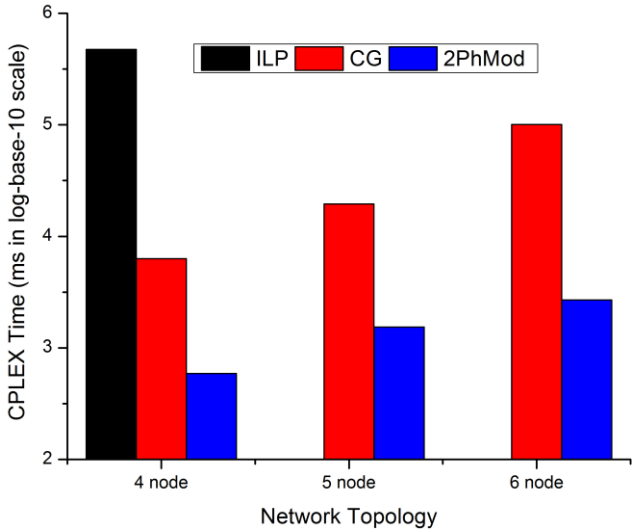


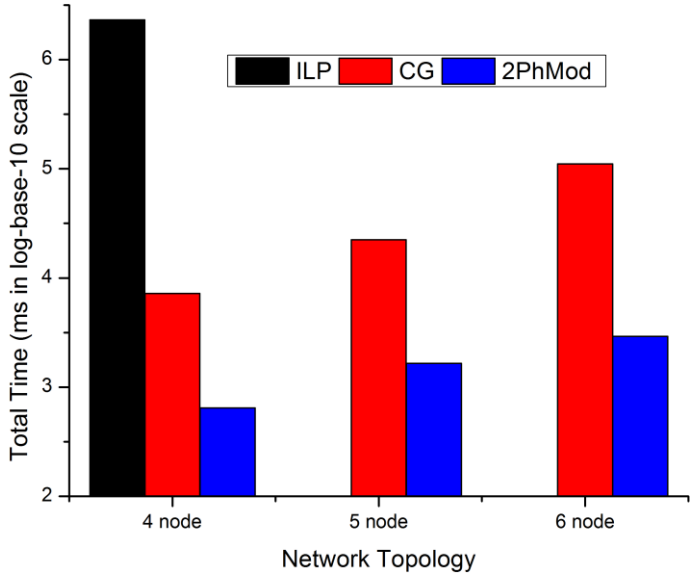
Continued…


1
2
4
3

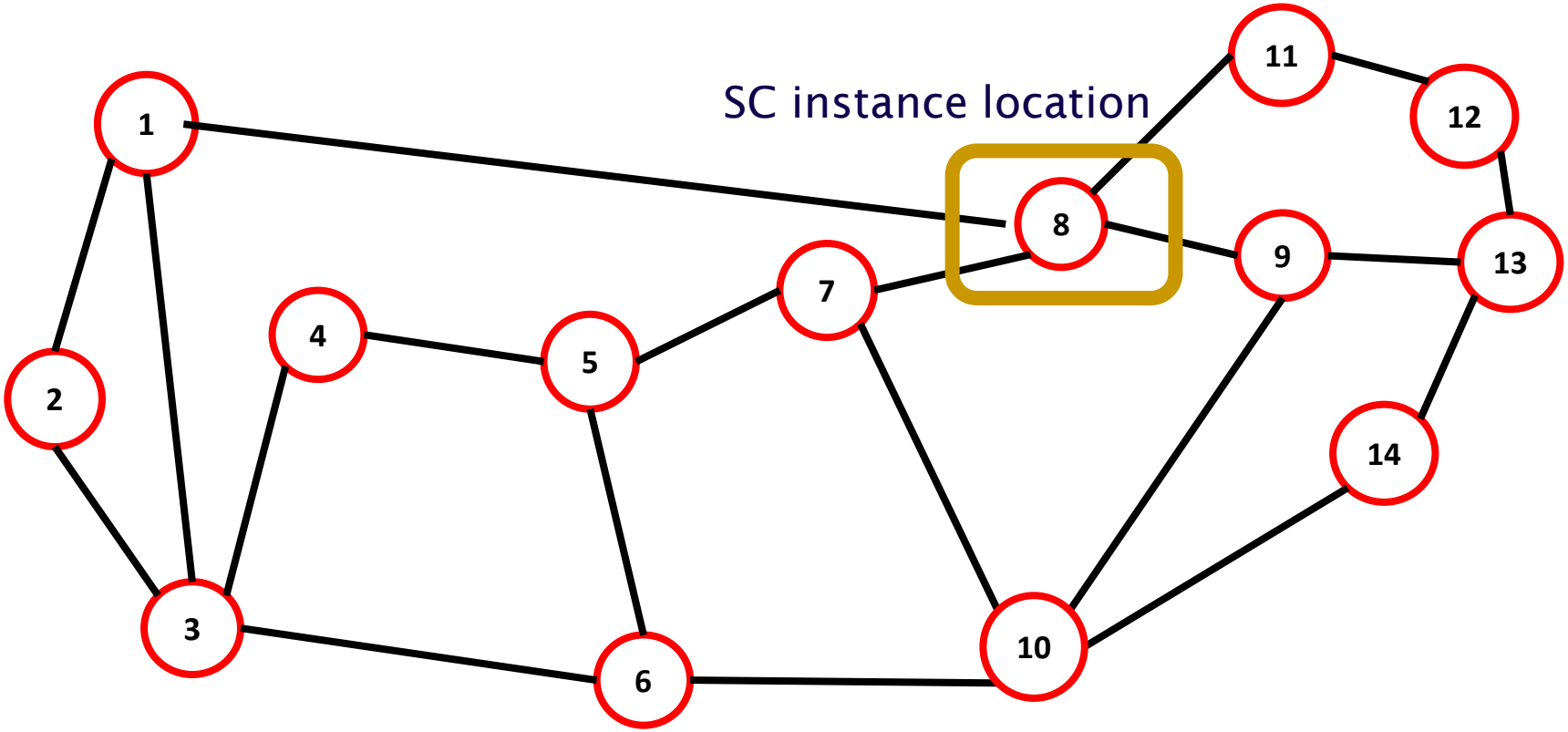

1
2
3
5
4


1
2
5
6
3
4


CPLEX Time (ms in log-base-10 scale)
6
5
4
3
2
ILP
CG
2PhMod
4 node
5 node
6 node
Network Topology

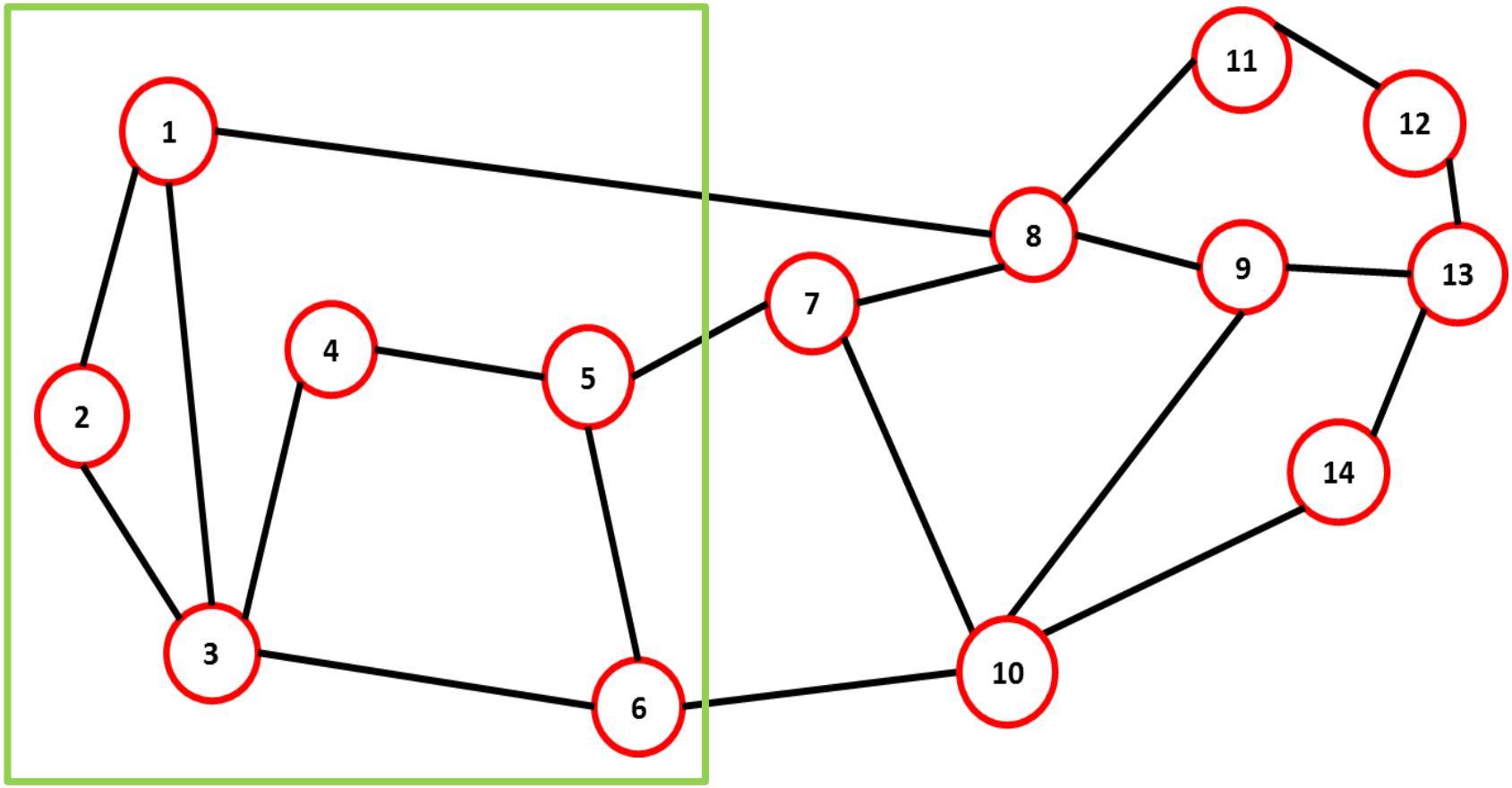

Total Time (ms in log-base-10 scale)
6
5
4
3
2
ILP
CG
2PhMod
4 node
5 node
6 node
Network Topology

Full Traffic Matrix, 1 SC deployment, 1 SC instance



All nodes are NFV-capable.
All node pairs have requests for the same service chain.

Grouping of traffic pairs

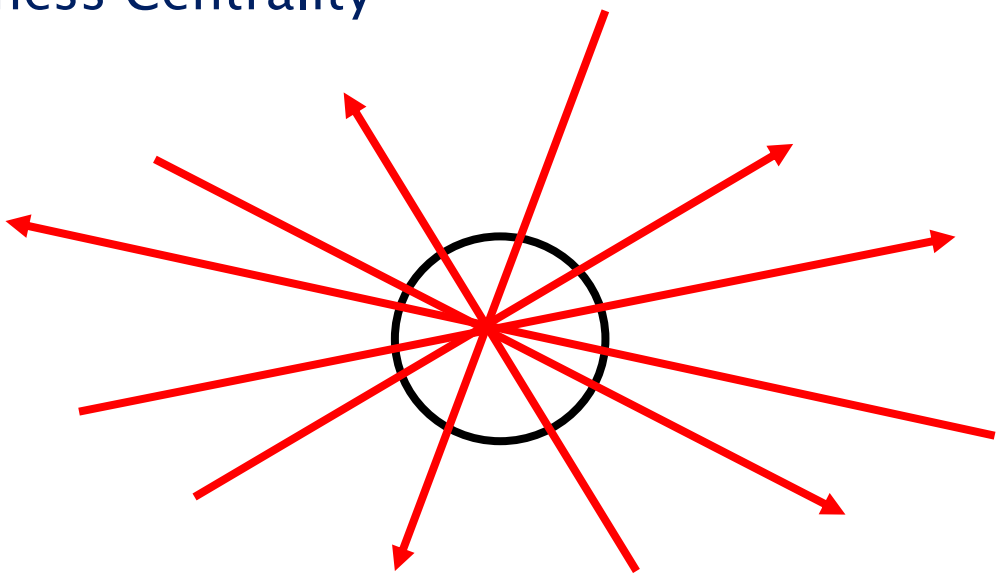


Continued …

- Create traffic flow groups
- Assign dummy SC Id's to traffic flow groups
- Big Question: **How to do we make traffic groups?**
- Model accounting for traffic groups becomes quadratic. Subsequent, linearization reduced the scalability of the model
- We, therefore, use a heuristic to do make the traffic groups

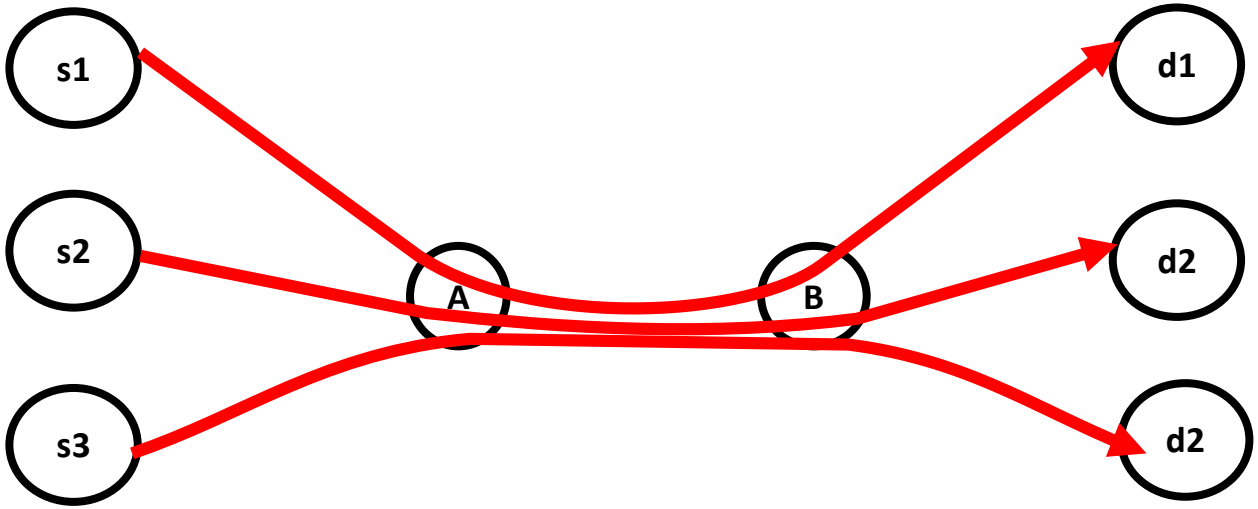
Grouping traffic flows around a node

- Betweenness Centrality



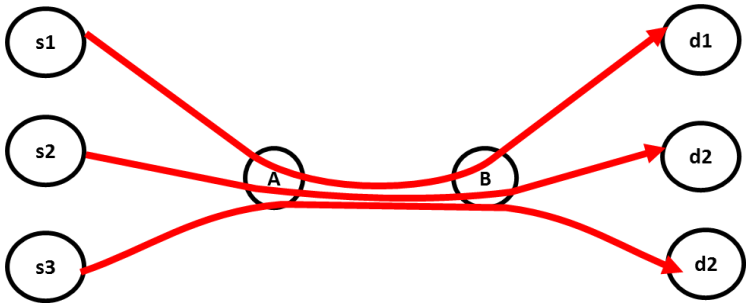
Group around node pairs of the graph

- A and B can also be source and destination
- Done for each SC



Continued…

- Ordered node pair with highest traffic flow count passing through on shortest paths



- Traffic flows which share sub-paths in common

- Deploying one SC instance for each such group

Shortest Path Traffic Grouping (SPTG)

- Given: the number of instances for a SC, the traffic flows for the SC

- The heuristic will:
  1. Find the **node pair** with highest number of flows
  2. This becomes another (s, d) pair **group/cluster**
  3. All flows in **group/cluster** are removed from global flow list
  4. Repeat step 1 to 3 until number of instances is reached
  5. Iterate through the remaining flows:
     1. Find best group based on which path length through **node pair**
     2. Add flow to that **group/cluster**

Shortest Path Traffic Grouping – Traffic Aware (SPTG-TA)

- Cluster around the **heaviest/largest** traffic flow
- The heuristic will:
  1. Find the **node pair** with highest number of flows **and the largest flow**
  2. This becomes another (s, d) pair **group/cluster**
  3. All flows in **group/cluster** are removed from global flow list
  4. Repeat step 1 to 3 until number of instances is reached
  5. Iterate through the remaining flows:
     1. Find best group based on which path length through **node pair**
     2. Add flow to that **group/cluster**

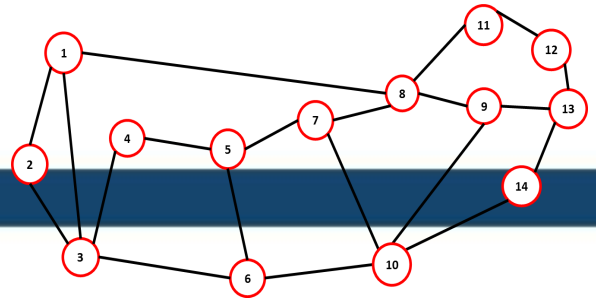
2 Phase Model

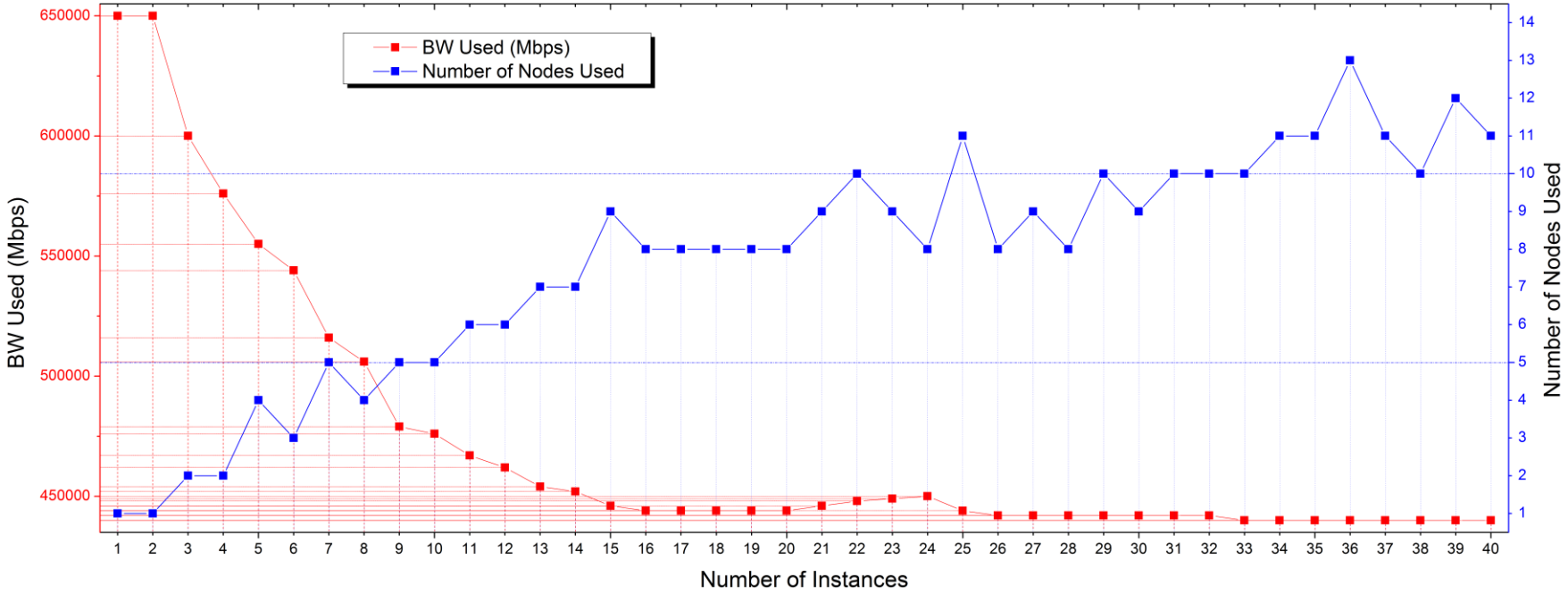
- 1st phase
  - Apply SPTG/SPTG-TA for each SC and create the required number of groups
  - Assign dummy SC ids to groups of flows
- 2nd phase
  - Use the column generation model which decides on 1 SC instance per SC
  - ***Also we can control the number of nodes that can host VNFs, we refer to this number as 'K'***

Assumptions

- All nodes are capable of hosting VNFs
- No CPU constraints are enforced
- No link capacity constraints are enforced
- Only one SC instance per SC model
- All traffic pairs have 1Gb traffic flow

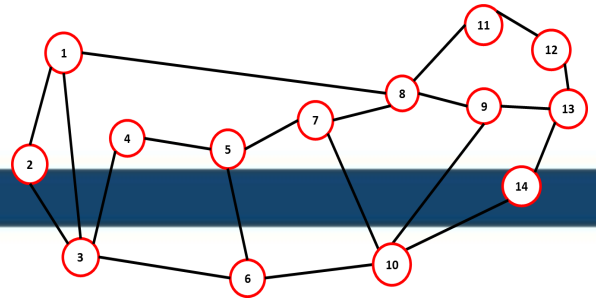
NSFNET K=14

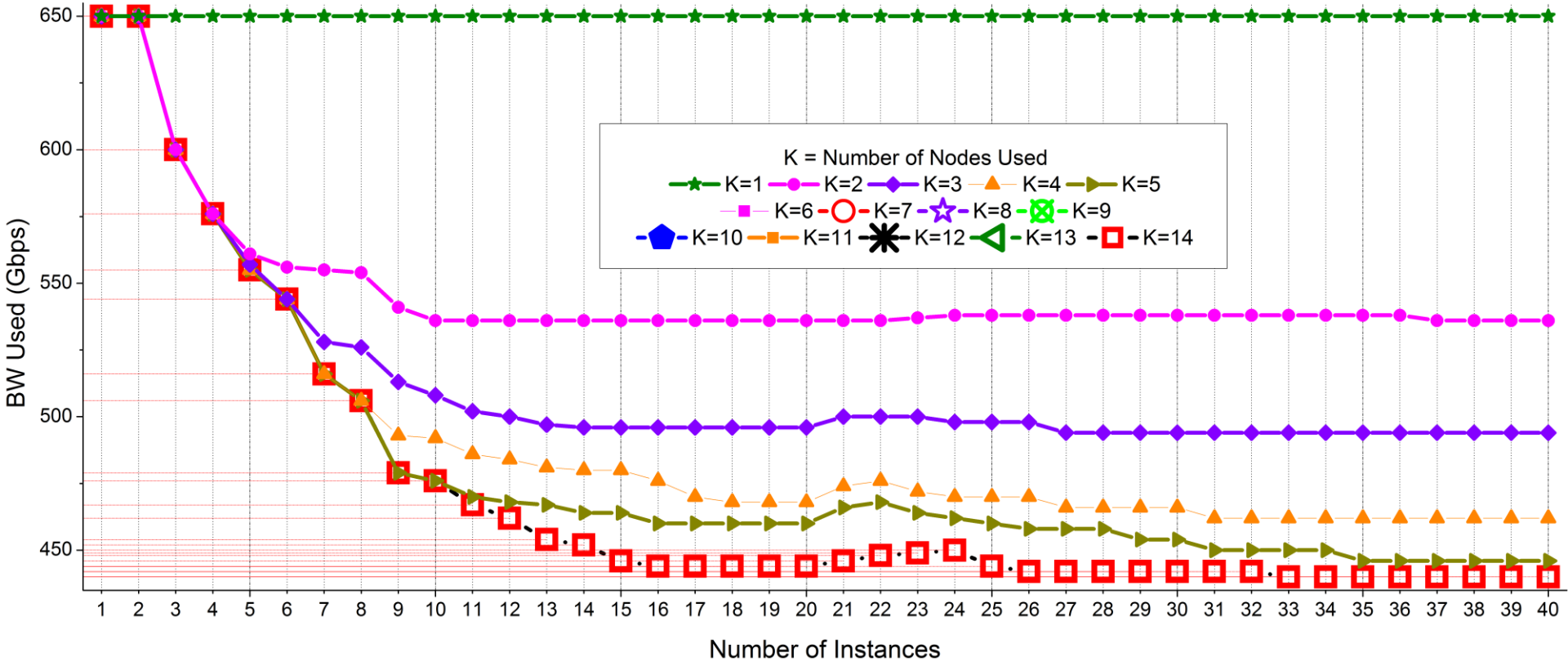




UCDAVIS
UNIVERSITY OF CALIFORNIA

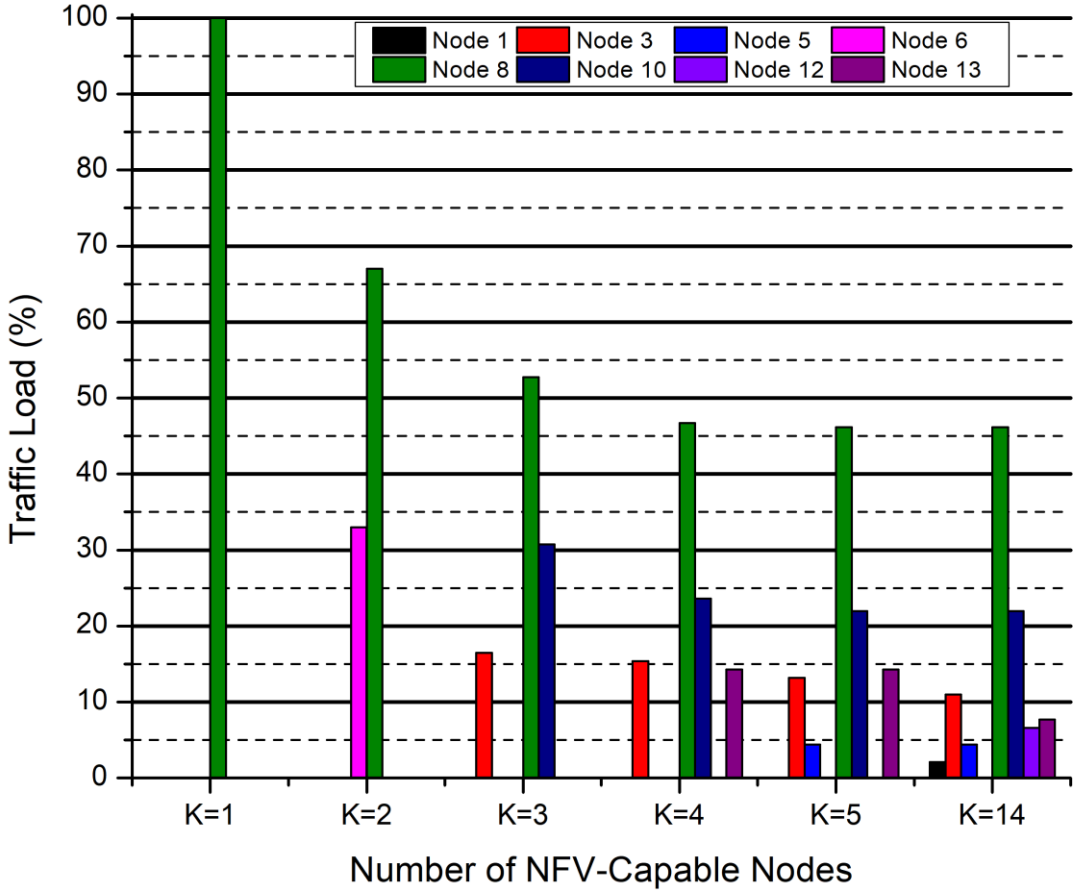
NSFNET K=14,5,4,3,2,1

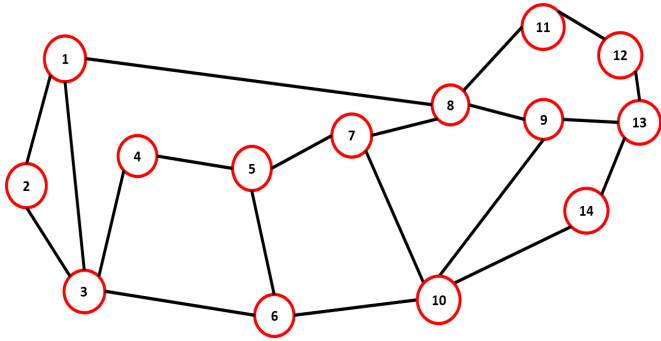




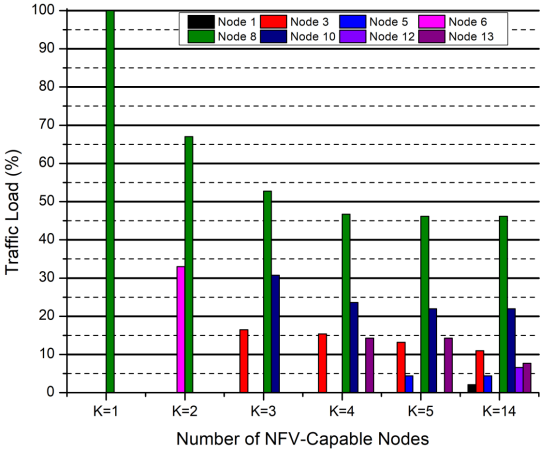
UCDAVIS
UNIVERSITY OF CALIFORNIA

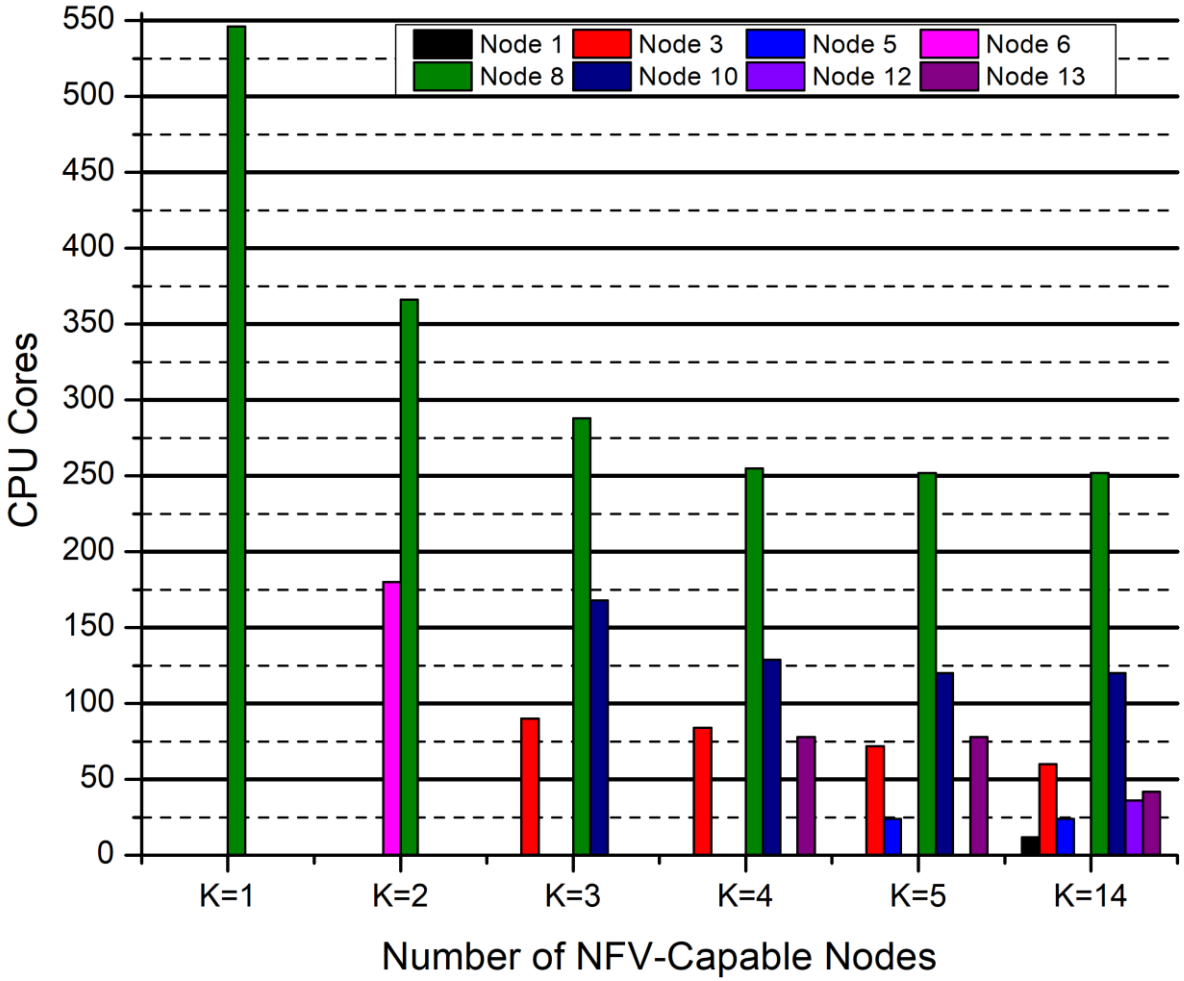
Partitioning of Traffic Load

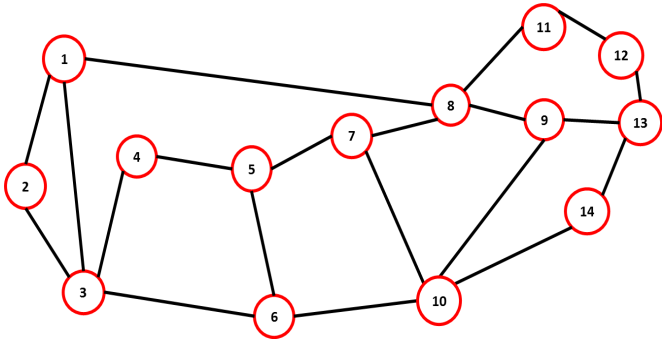




Partitioning of CPU resources

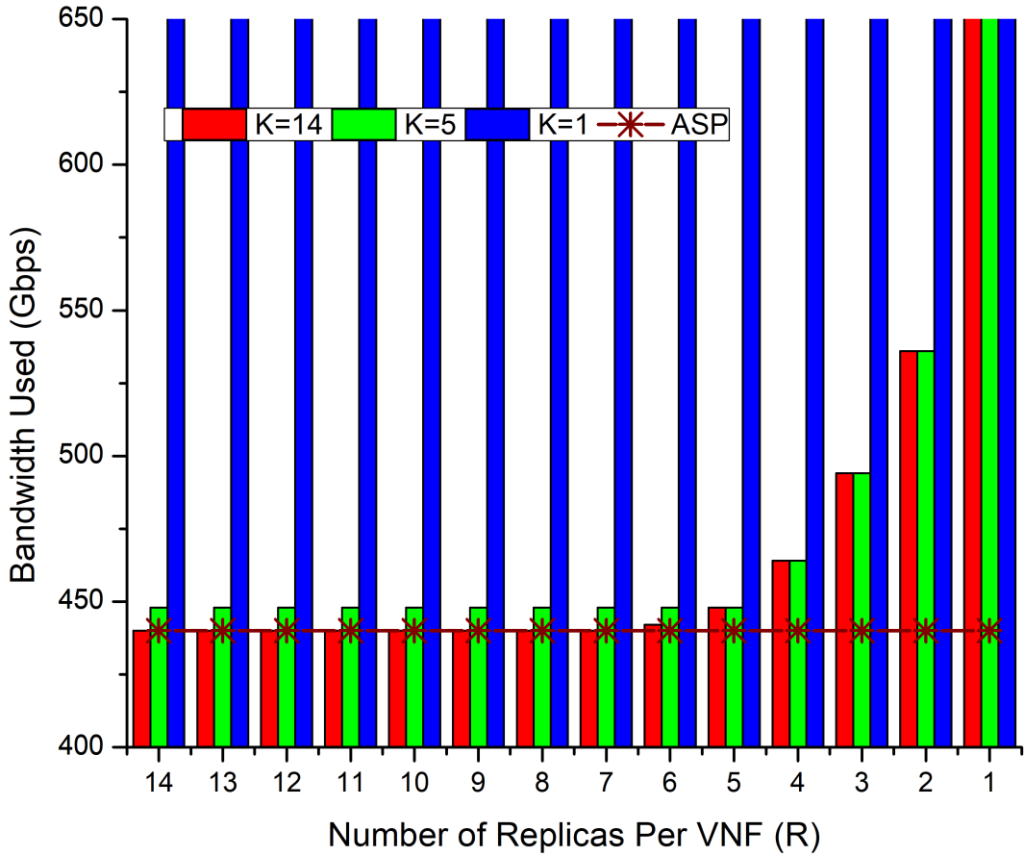




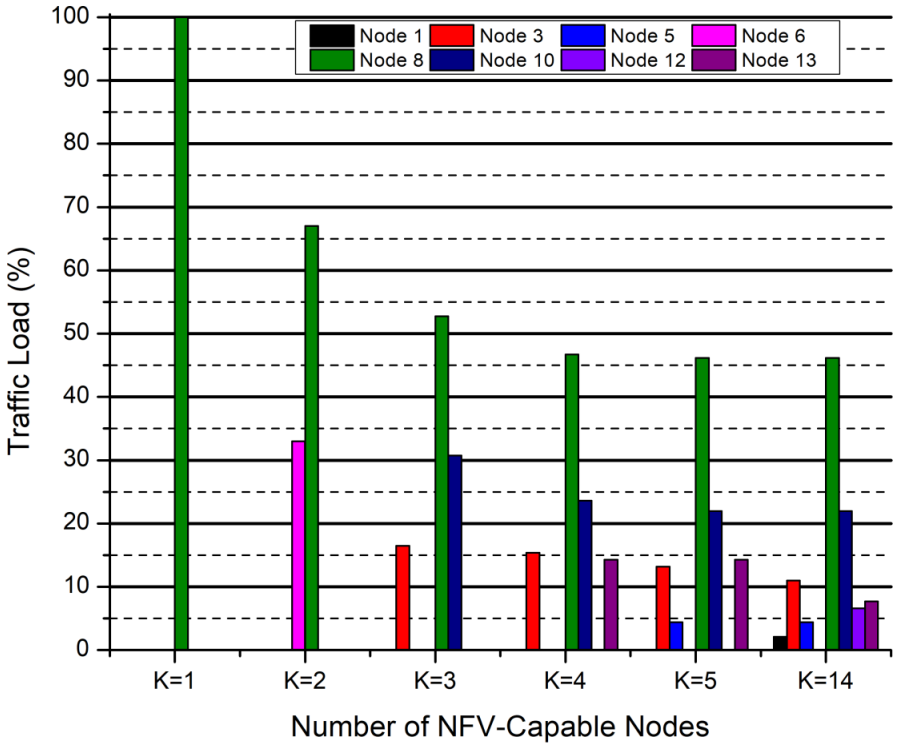


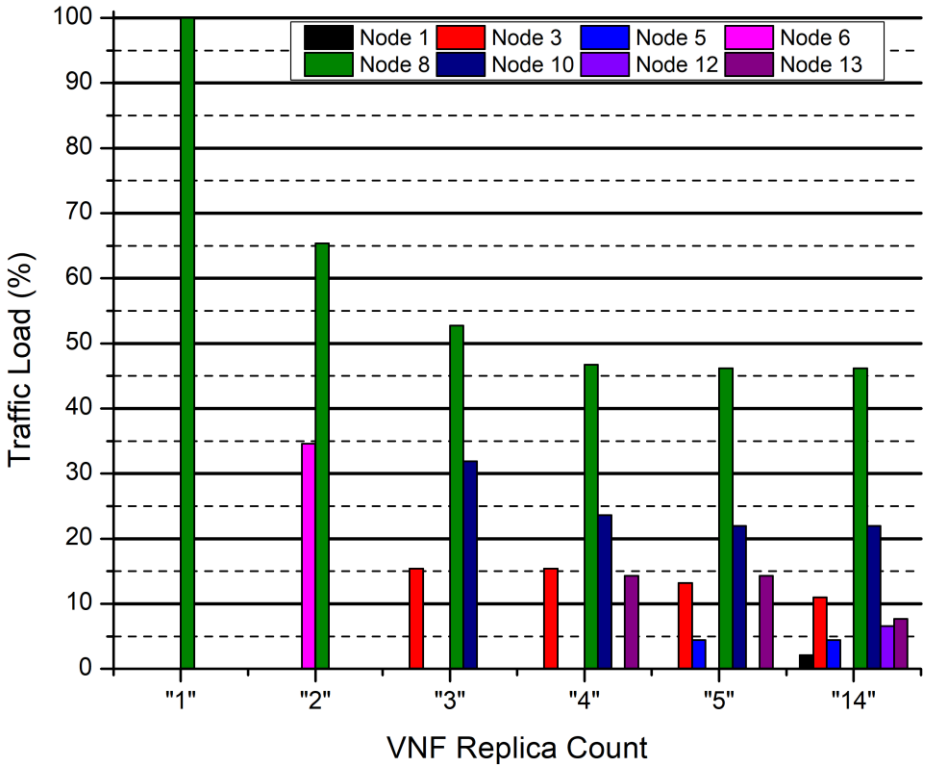
UCDAVIS
UNIVERSITY OF CALIFORNIA

VNF Replica Constraints



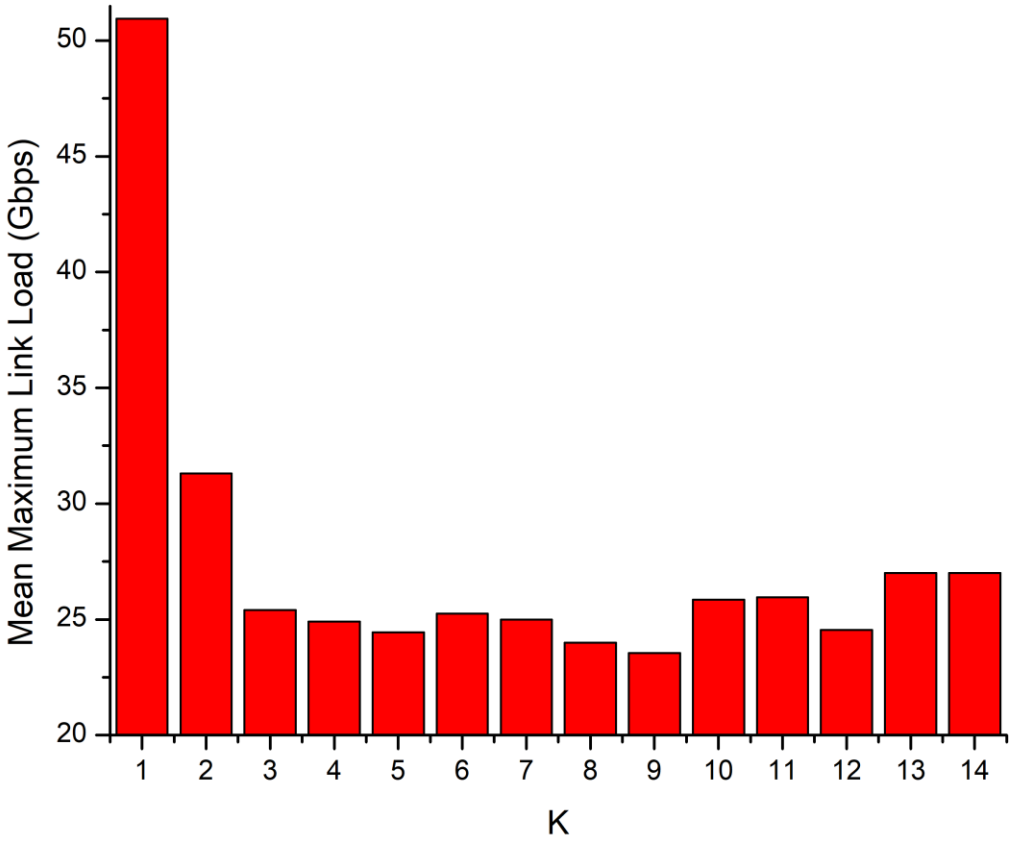
Correlation between ‘K’ and ‘VNF Replica Count’



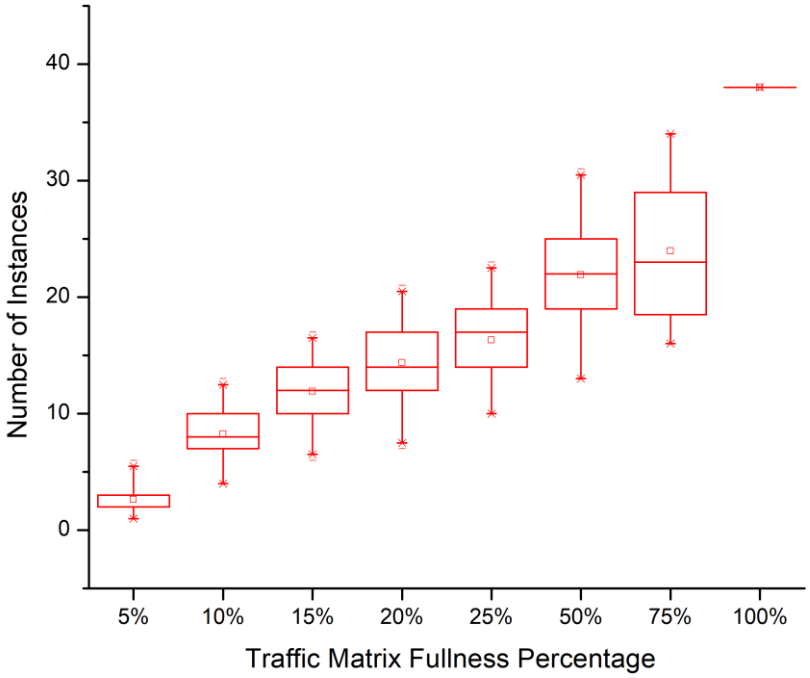


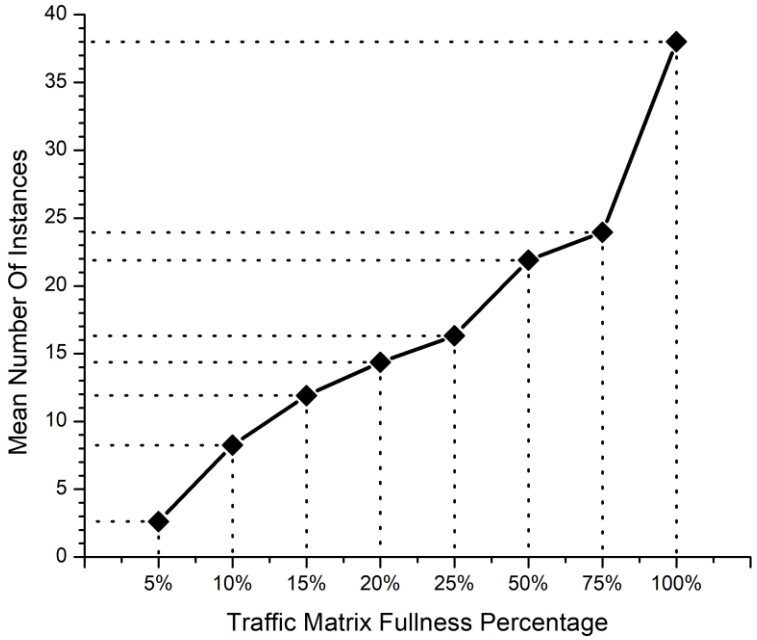
UCDAVIS
UNIVERSITY OF CALIFORNIA

Mean Maximum Link Load



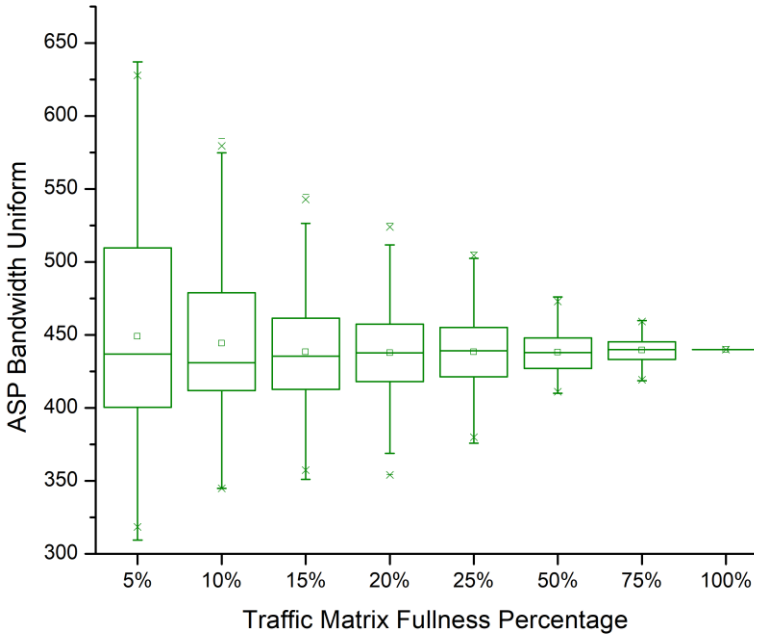
Cluster Counts (Uniform Traffic – Same Load)

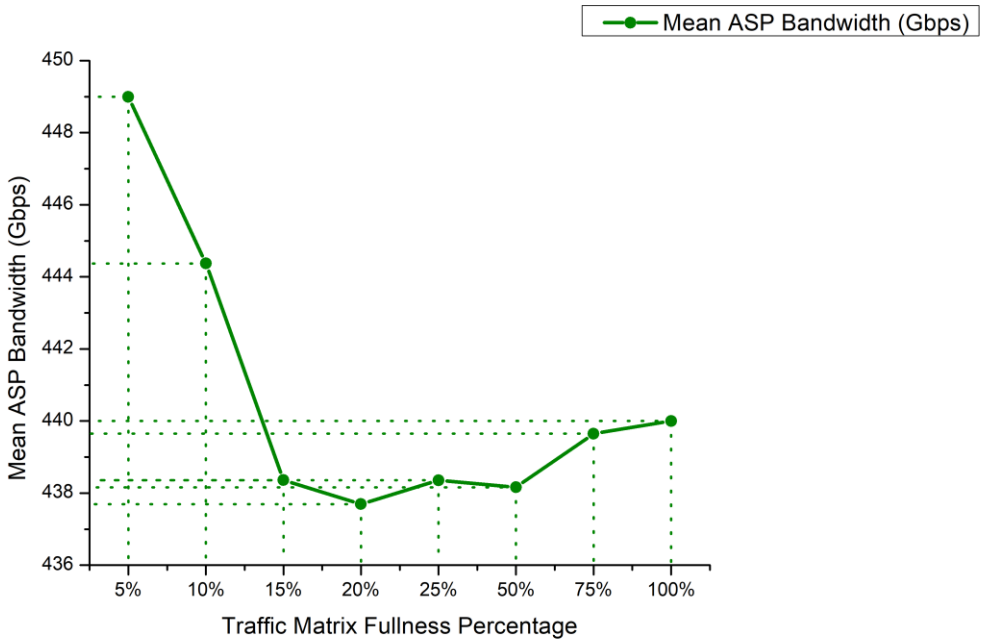




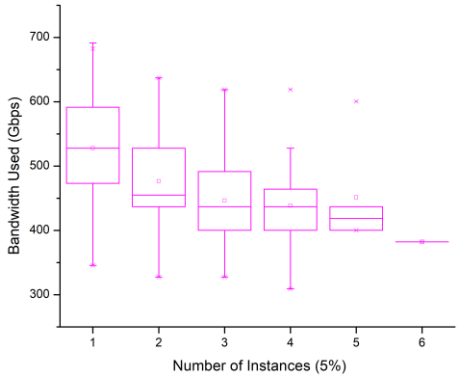
UCDAVIS
UNIVERSITY OF CALIFORNIA

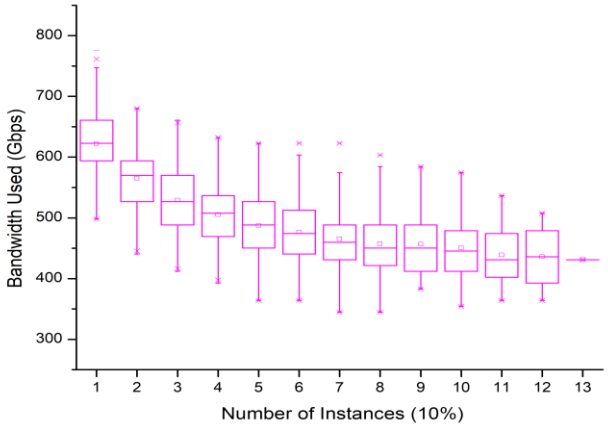
ASP Bandwidth Variation (Uniform Traffic – Same Load)

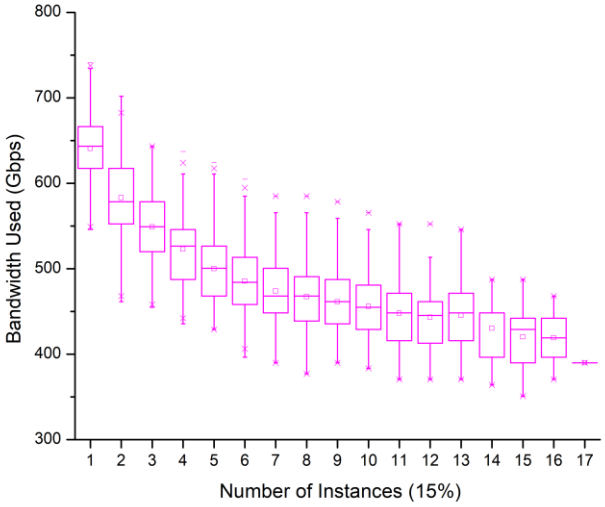


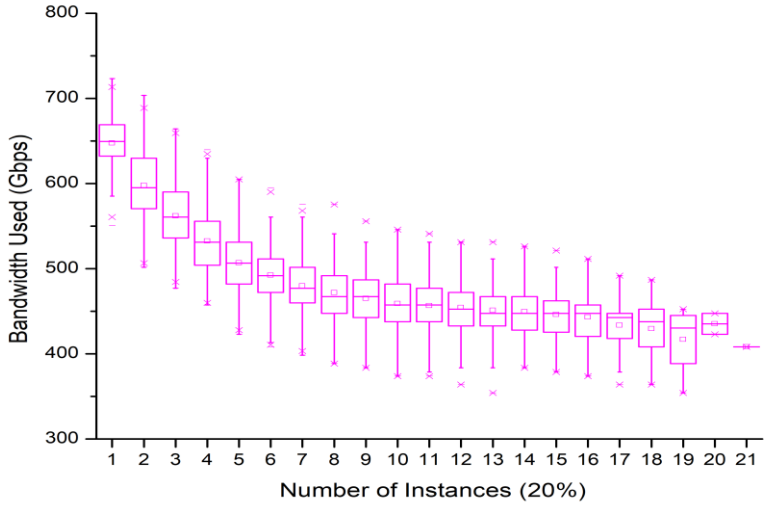


Bandwidth Used (Uniform, Same Load, Across Clusters and Traffic)

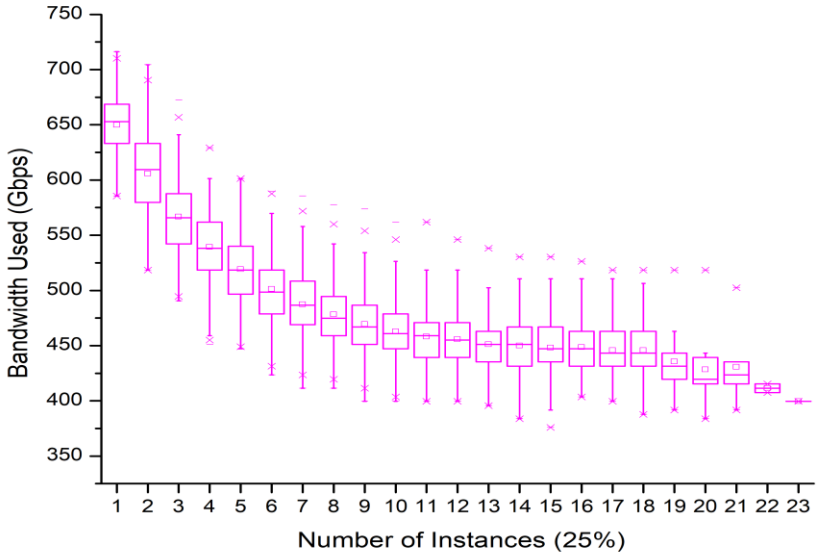


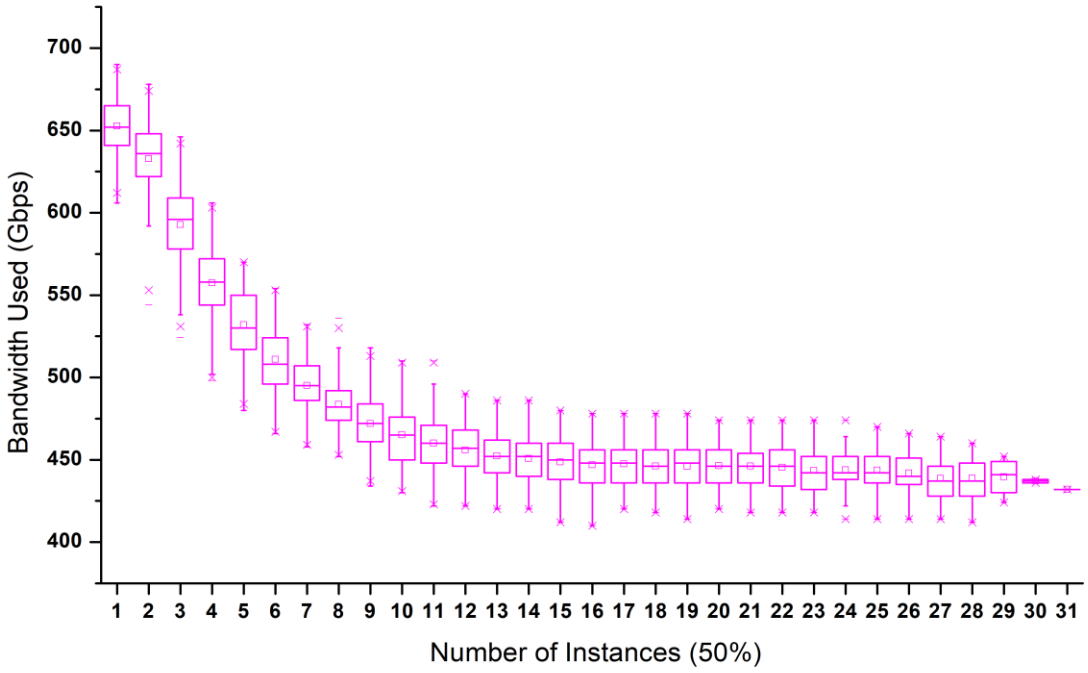




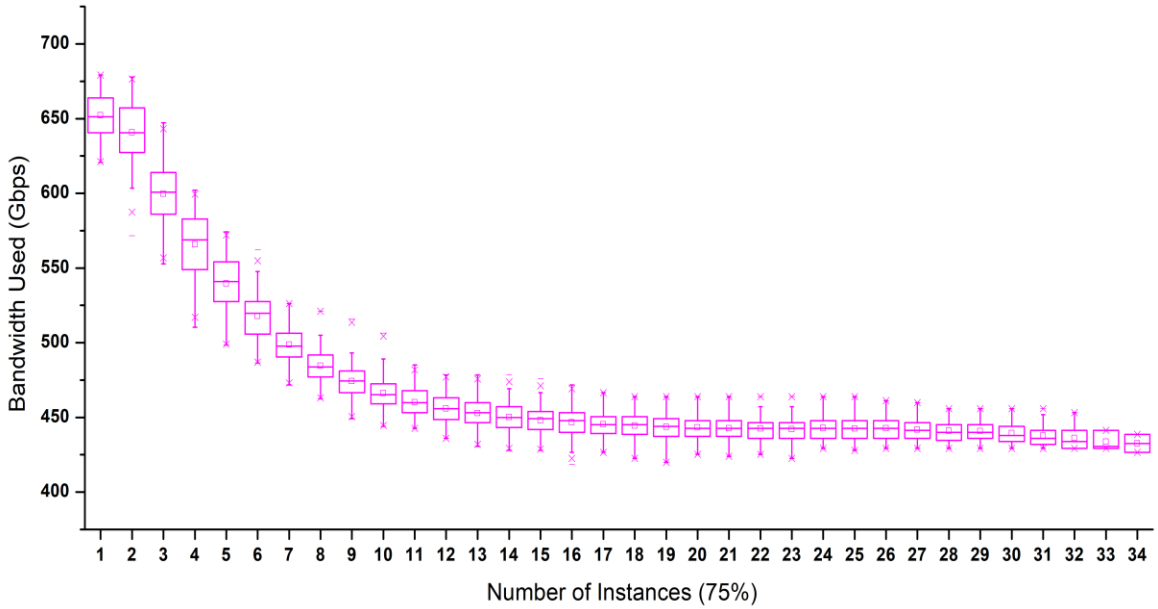


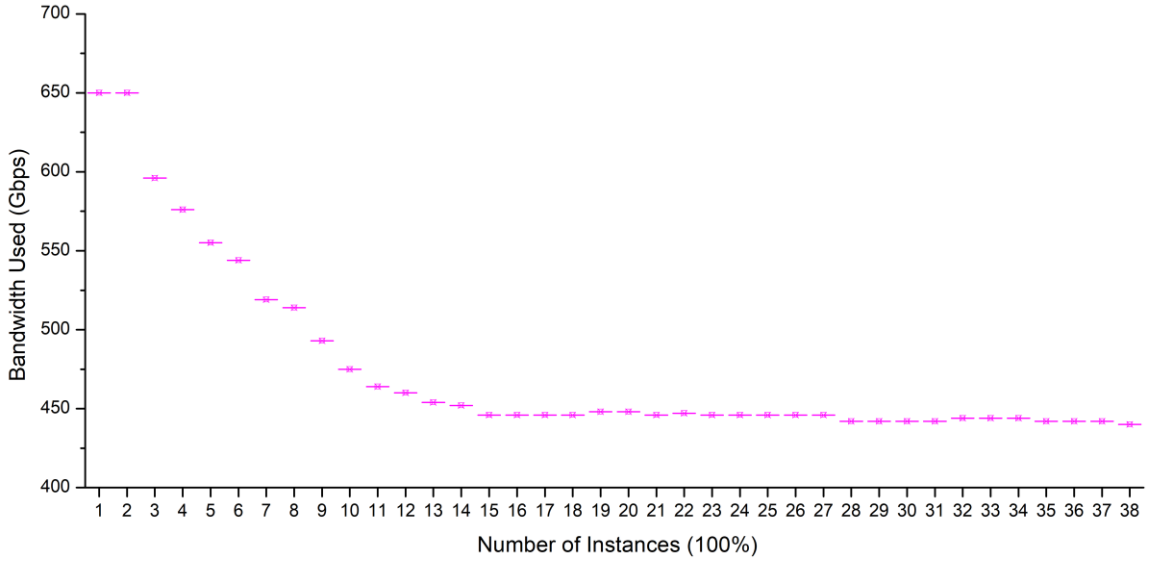
Continued…


Bandwidth Used (Gbps)
750
700
650
600
550
500
450
400
350
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Number of Instances (25%)

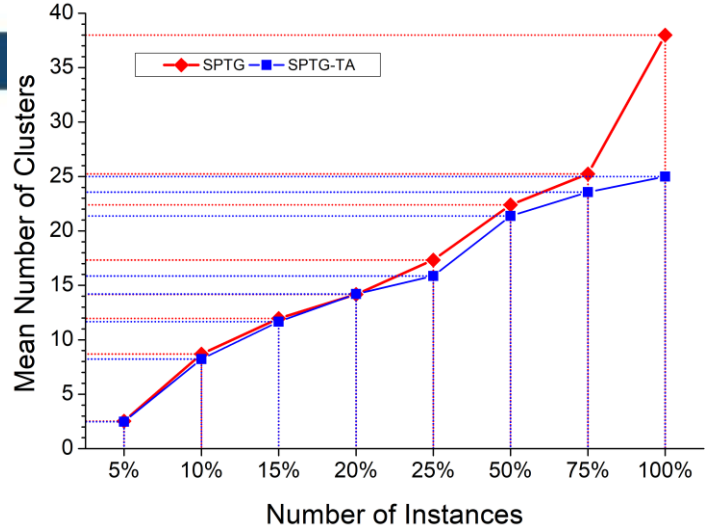

Bandwidth Used (Gbps)
700
650
600
550
500
450
400
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Number of Instances (50%)

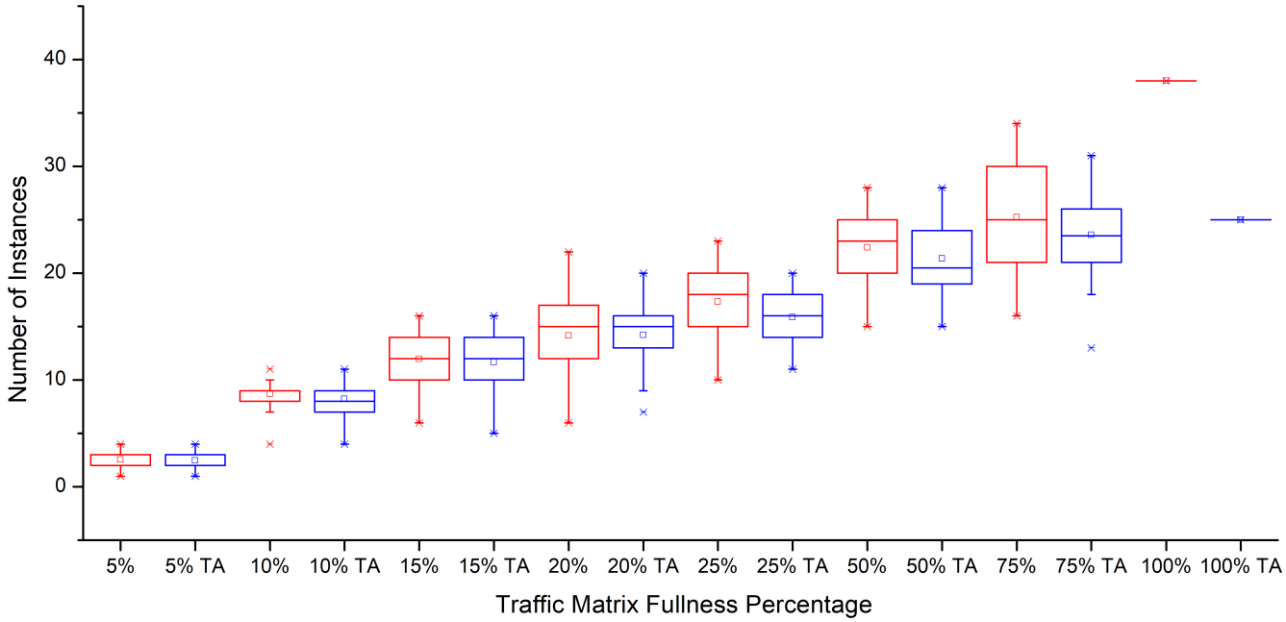
Continued…


Bandwidth Used (Gbps)
700
650
600
550
500
450
400
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
Number of Instances (75%)


Bandwidth Used (Gbps)
700
650
600
550
500
450
400
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Number of Instances (100%)

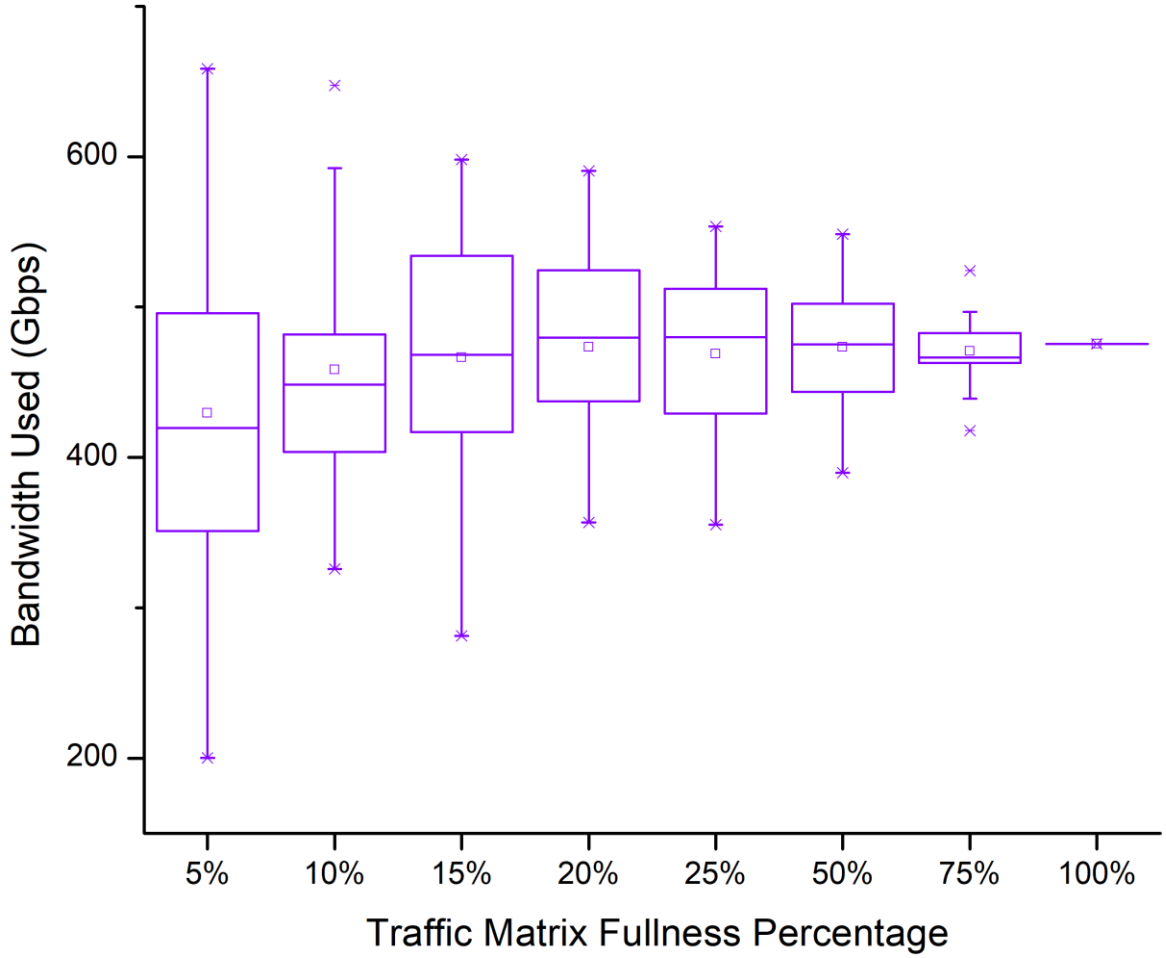
Compare Cluster Counts (Skewed Traffic – Same Load)





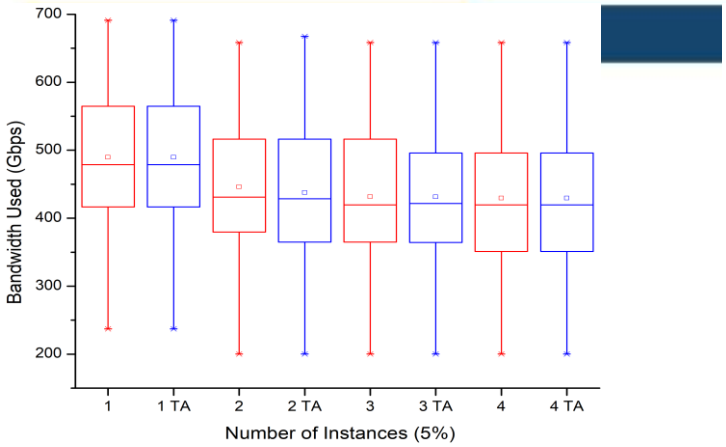
UCDAVIS
UNIVERSITY OF CALIFORNIA

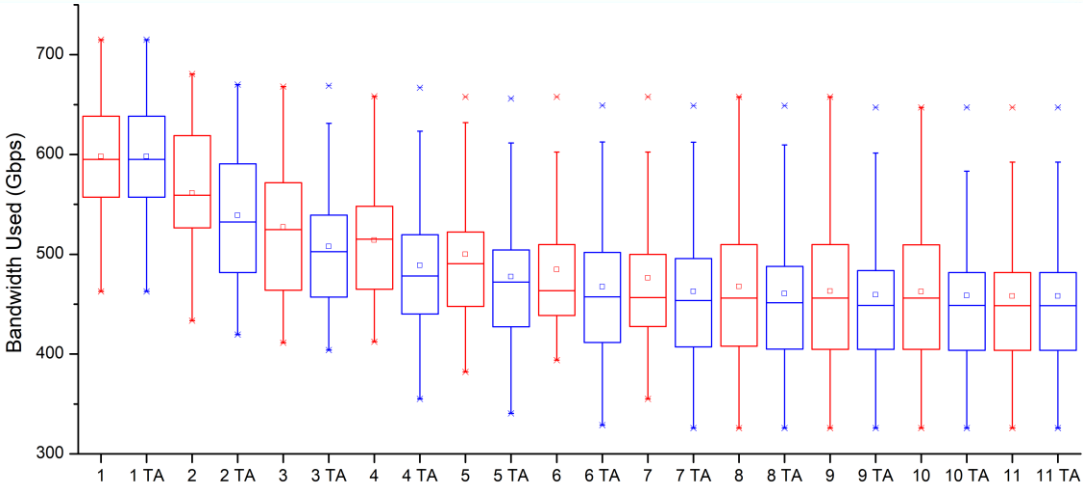
ASP Bandwidth Variation (Skewed Traffic – Same Load)

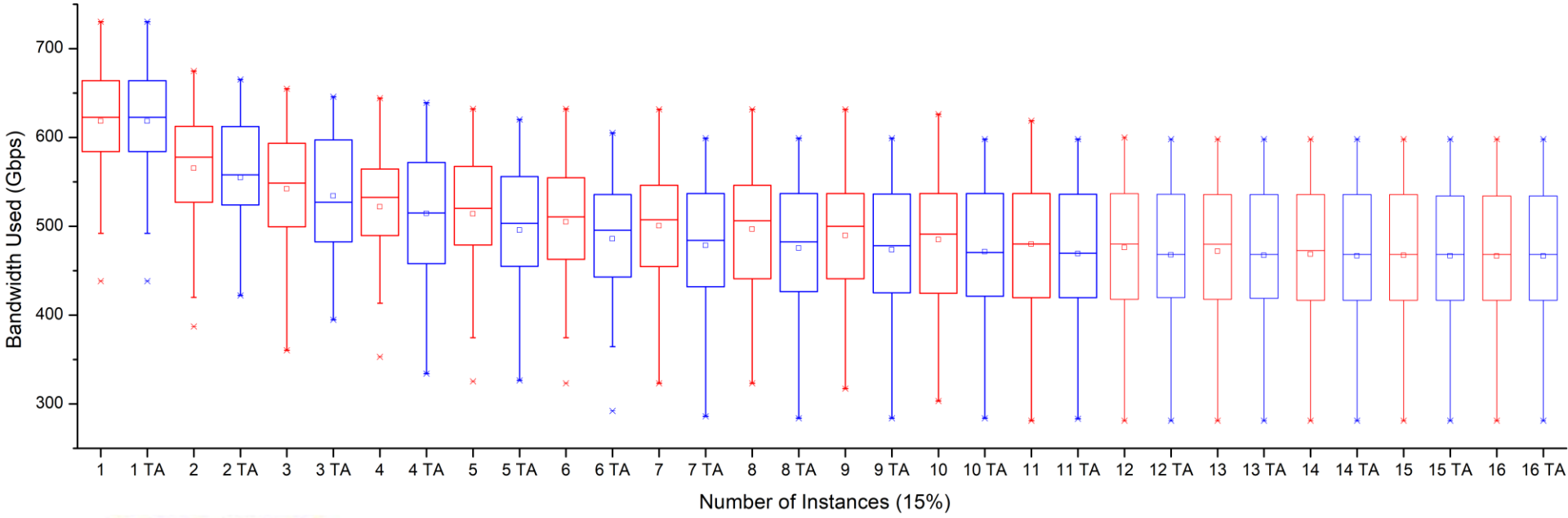


UCDAVIS
UNIVERSITY OF CALIFORNIA

Bandwidth Used (Skewed, Same Load, Across Clusters and Traffic)

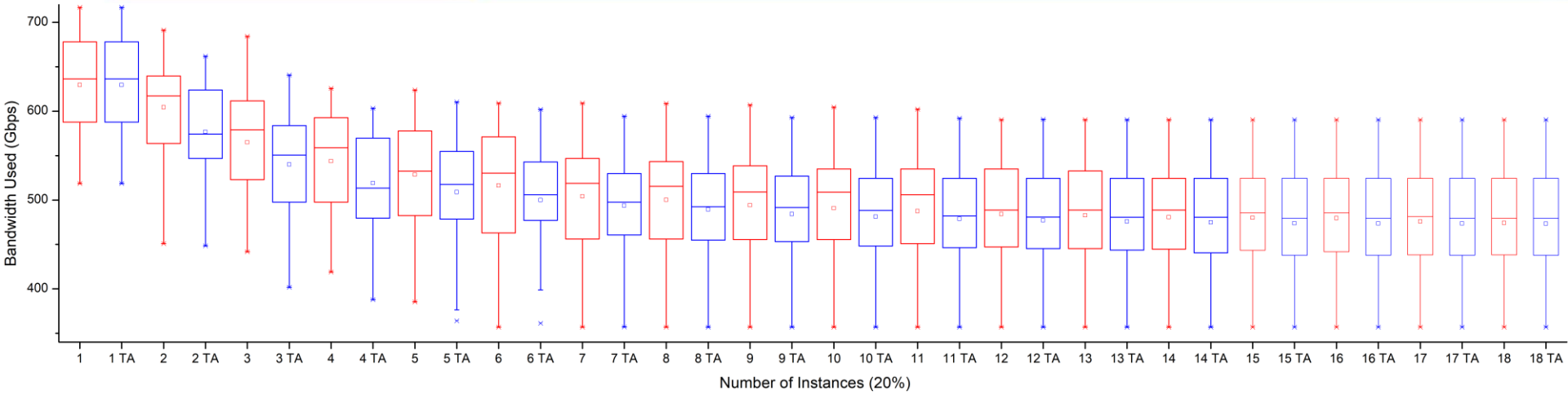


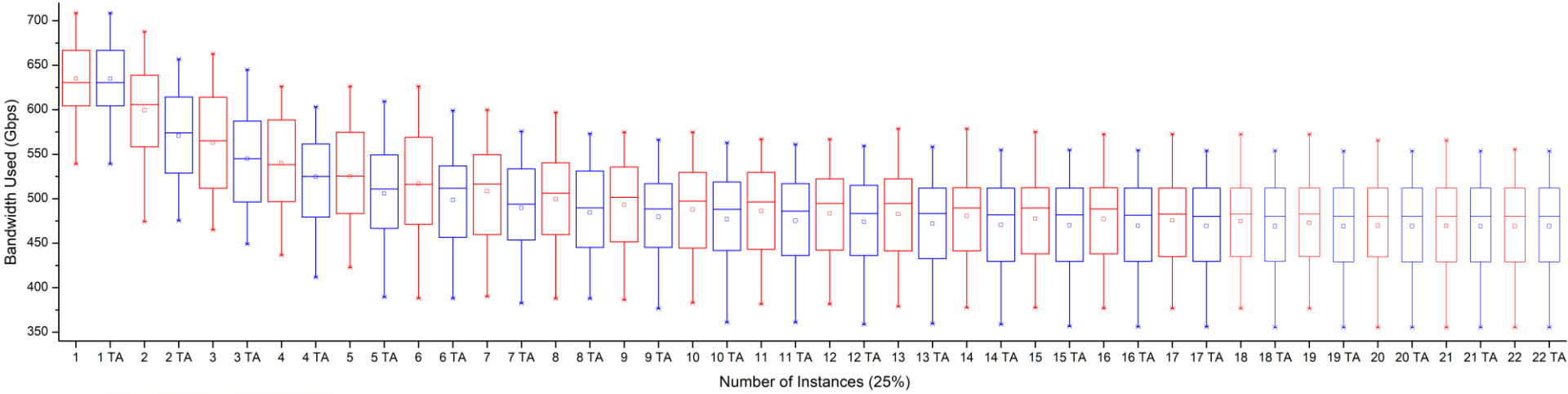




UC DAVIS
UNIVERSITY OF CALIFORNIA

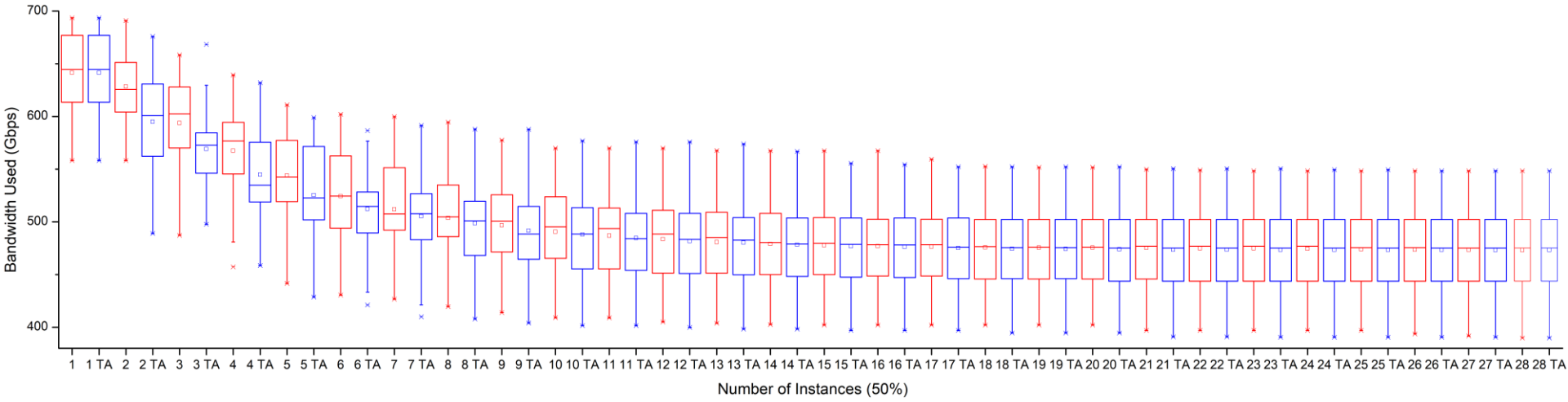
Continued…

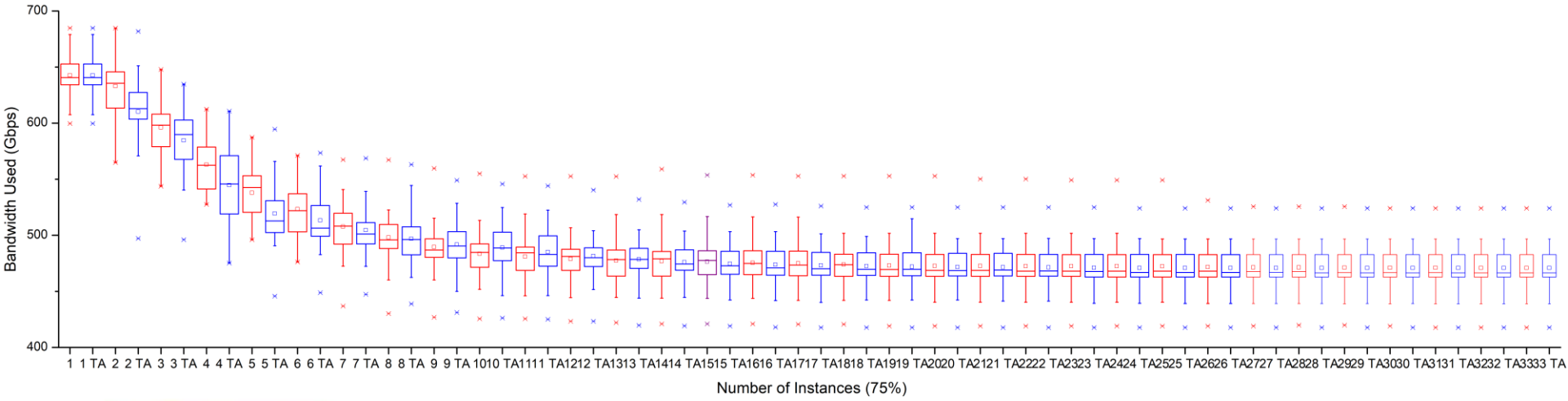




UC DAVIS
UNIVERSITY OF CALIFORNIA

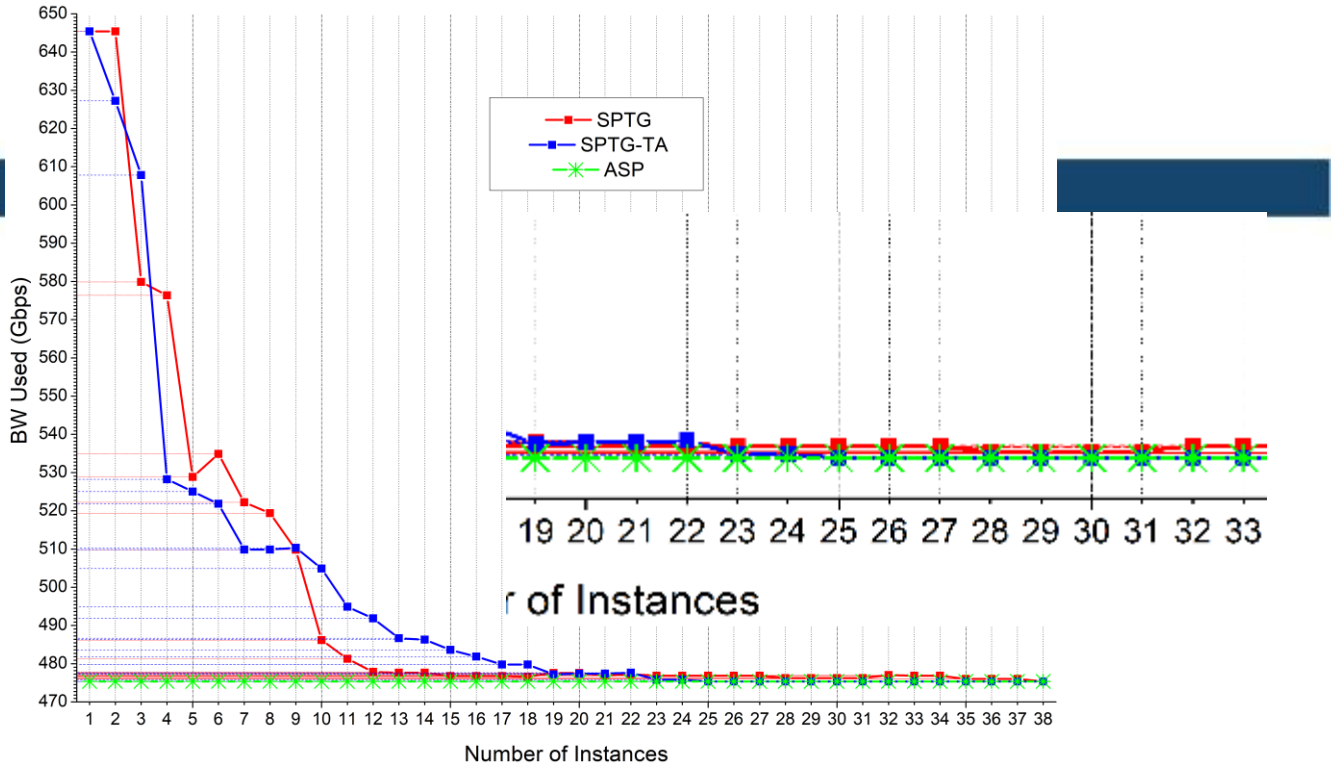
Continued…

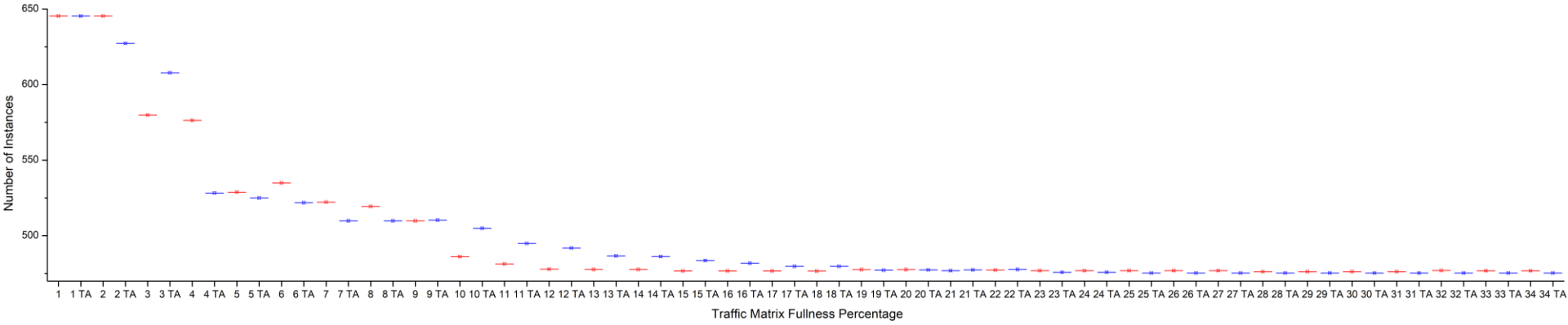




UC DAVIS
UNIVERSITY OF CALIFORNIA

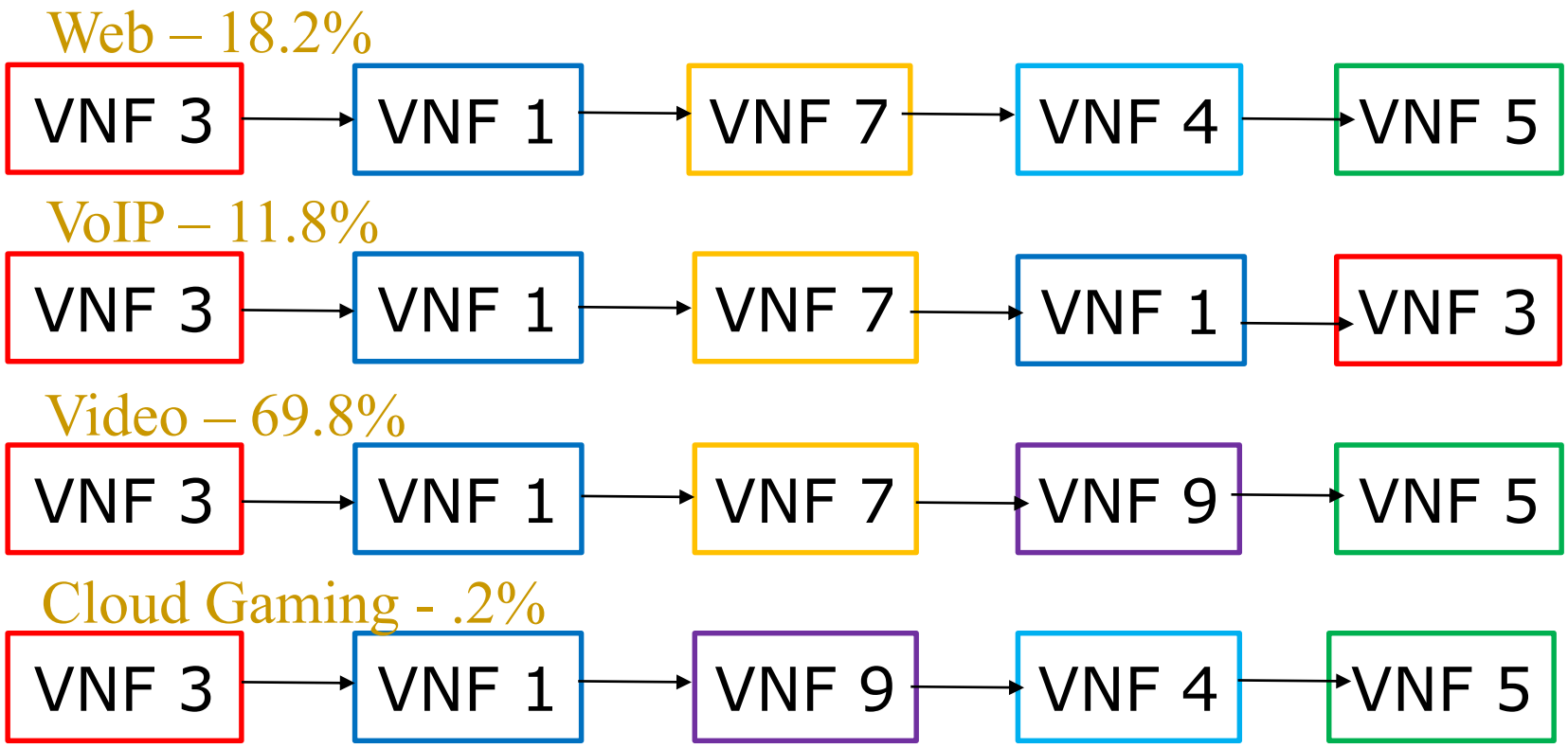
Continued…





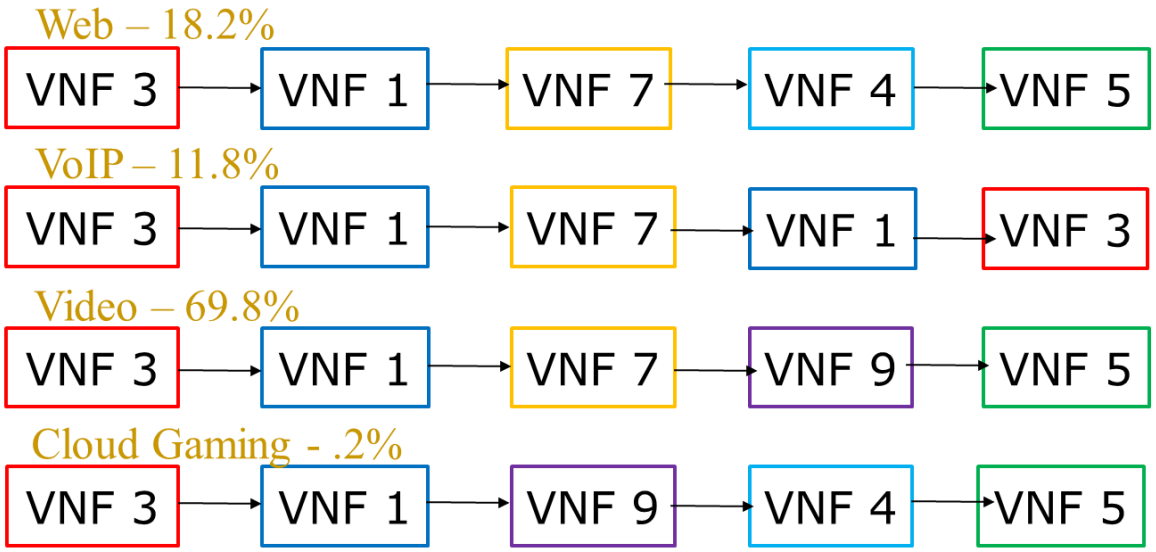
UC DAVIS
UNIVERSITY OF CALIFORNIA

Scenario 2 (4 Service Chains, 1 Tb Load)



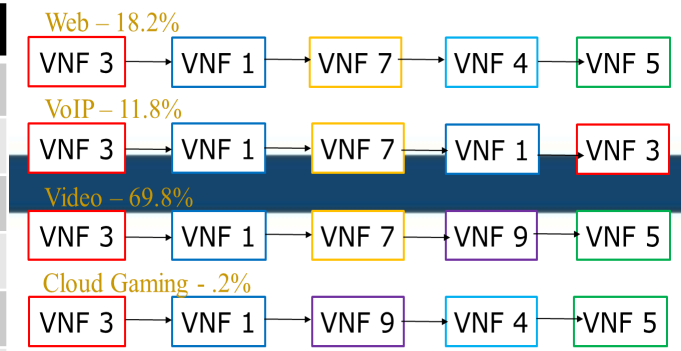
Continued…

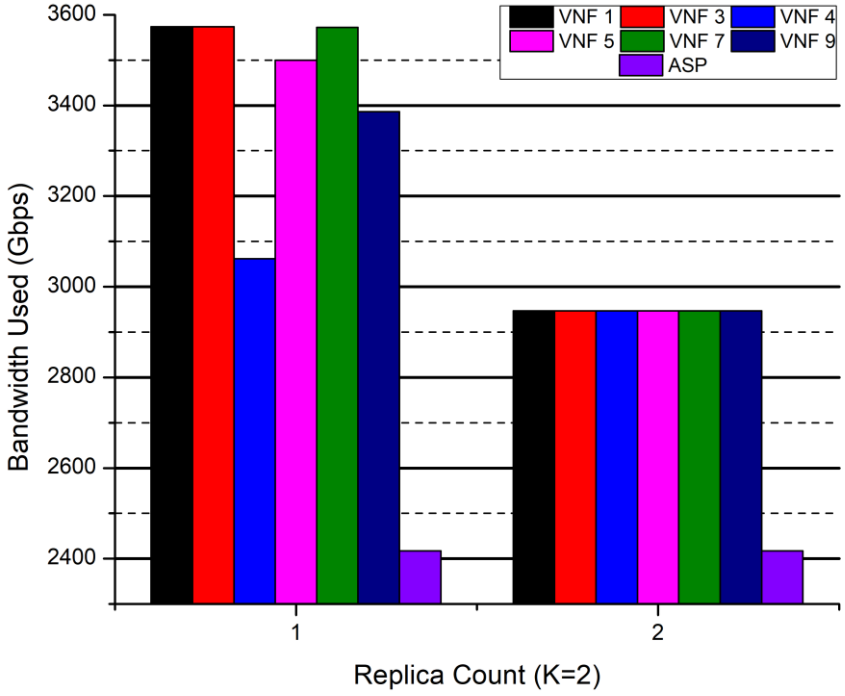
VNF ID	SC's	Traffic Load (Tbps)
3	0,1,2,3	1.118
1	0,1,2,3	1.118
7	0,1,2	.998
5	0,2,3	.882
4	0,3	.184
9	2,3	.700

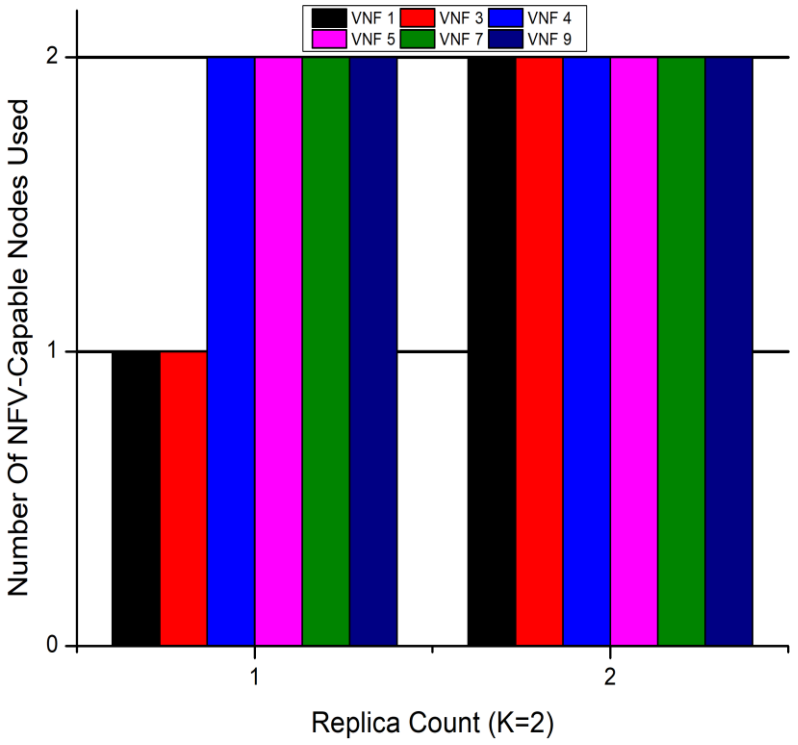


K = 2

VNF ID	SC's	Traffic Load (Tbps)
3	0,1,2,3	1.118
1	0,1,2,3	1.118
7	0,1,2	.998
5	0,2,3	.882
4	0,3	.184
9	2,3	.700

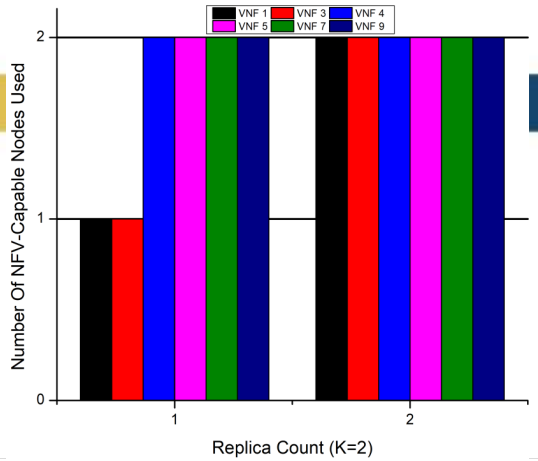




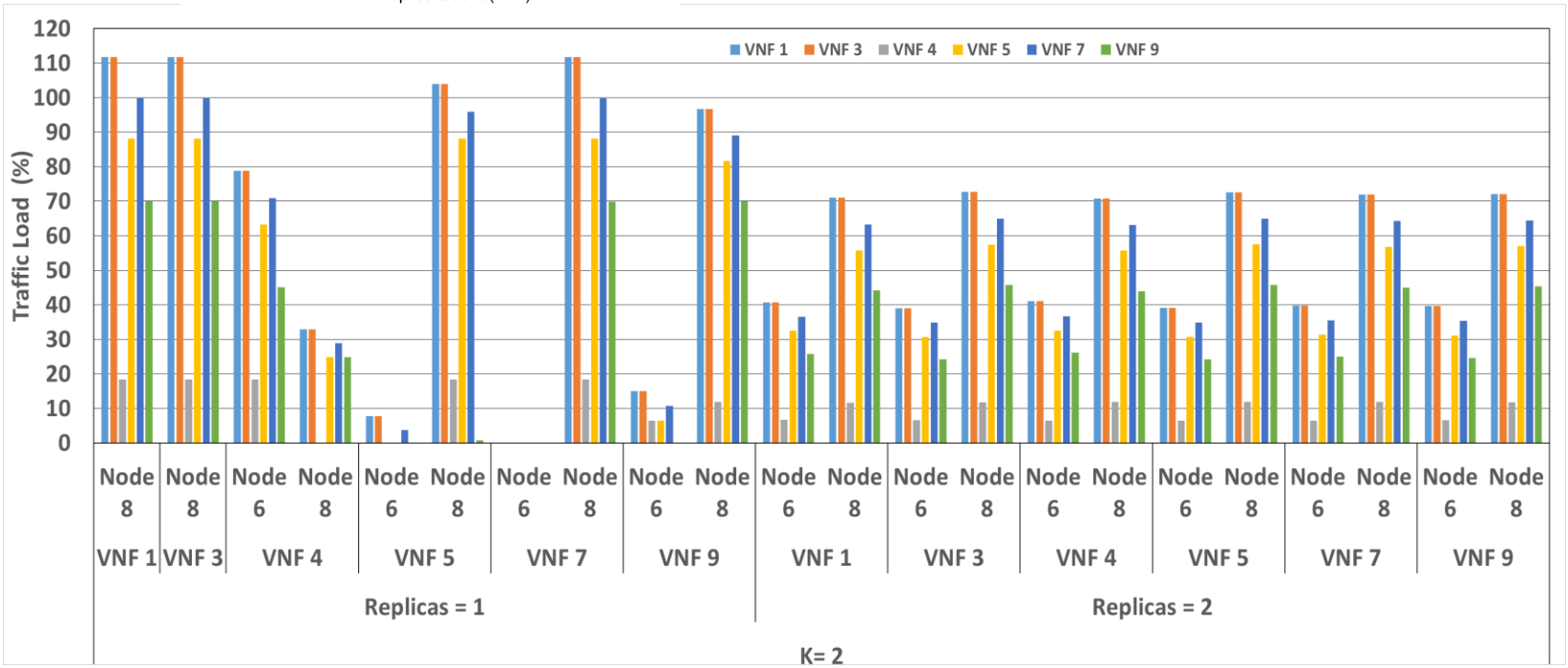


UC DAVIS
UNIVERSITY OF CALIFORNIA

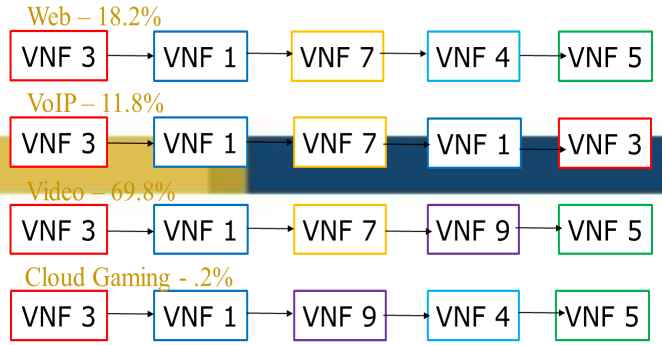
Continued…



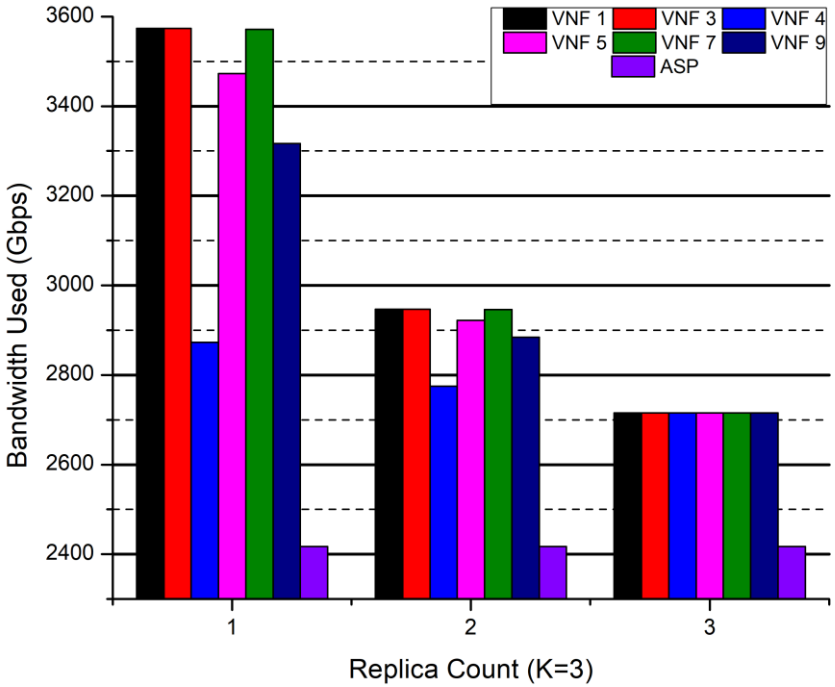
VNF ID	SC's	Traffic Load (Tbps)
3	0,1,2,3	1.118
1	0,1,2,3	1.118
7	0,1,2	.998
5	0,2,3	.882
4	0,3	.184
9	2,3	.700

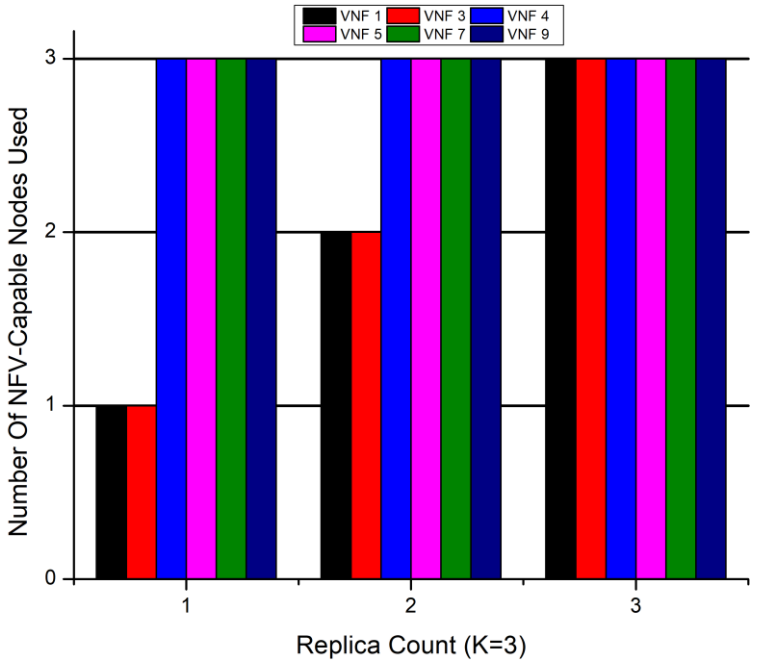


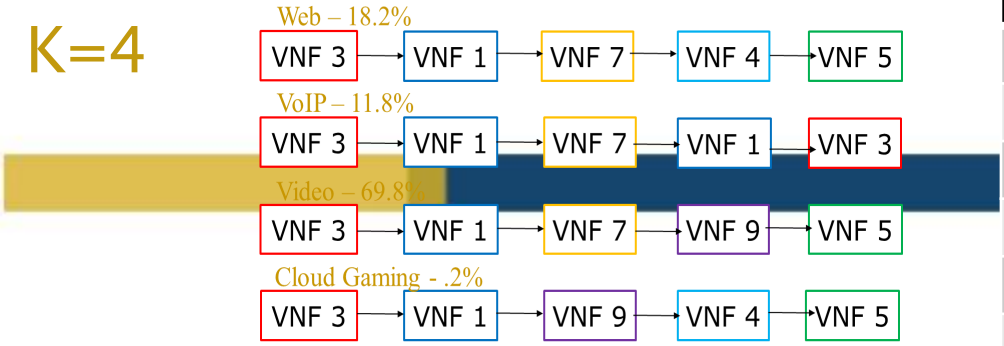
K=3



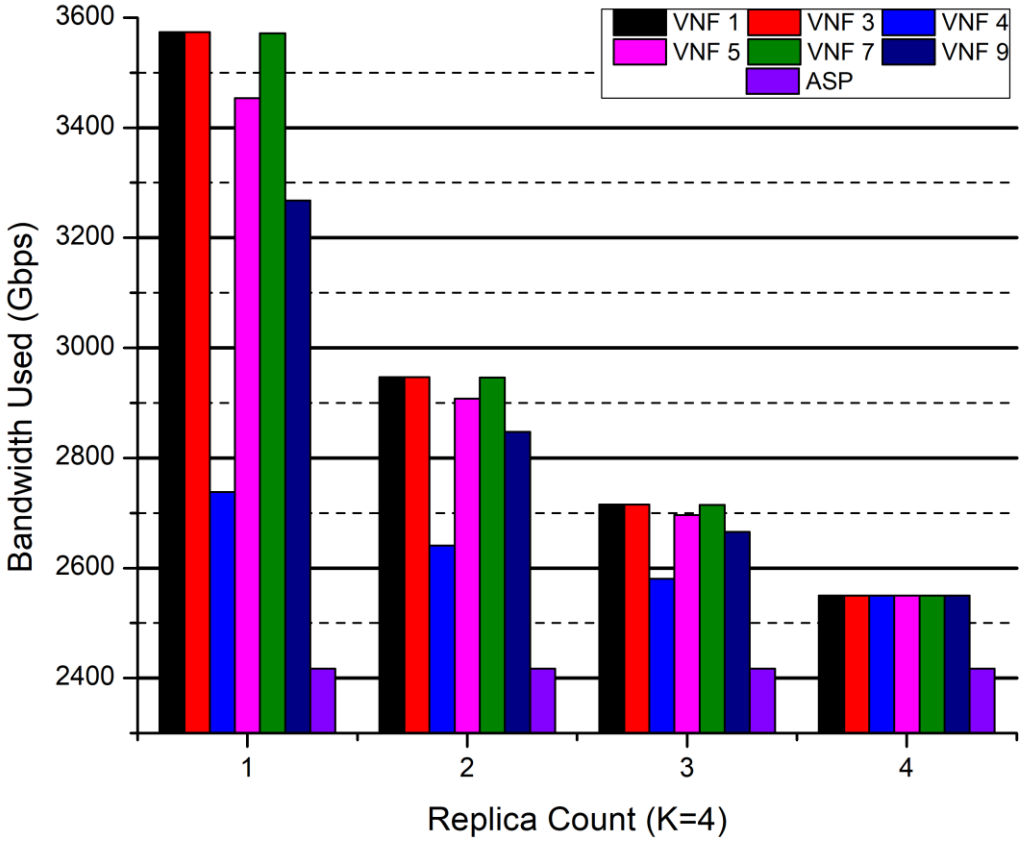
VNF ID	SC's	Traffic Load (Tbps)
3	0,1,2,3	1.118
1	0,1,2,3	1.118
7	0,1,2	.998
5	0,2,3	.882
4	0,3	.184
9	2,3	.700

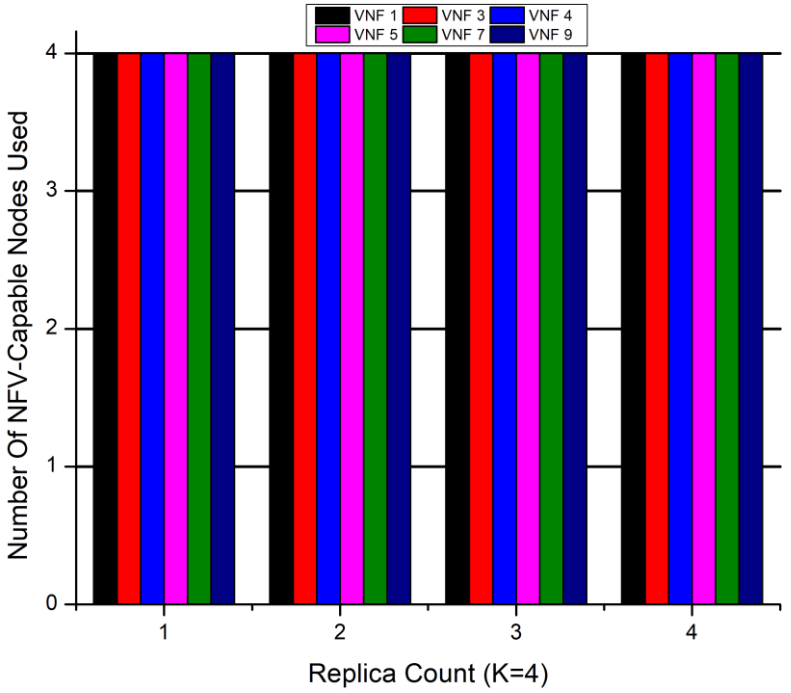




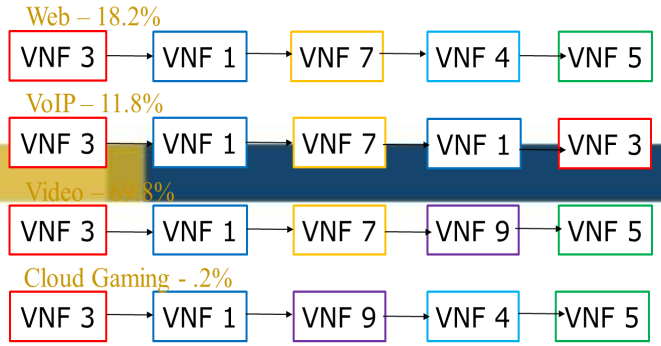


VNF ID	SC's	Traffic Load (Tbps)
3	0,1,2,3	1.118
1	0,1,2,3	1.118
7	0,1,2	.998
5	0,2,3	.882
4	0,3	.184
9	2,3	.700

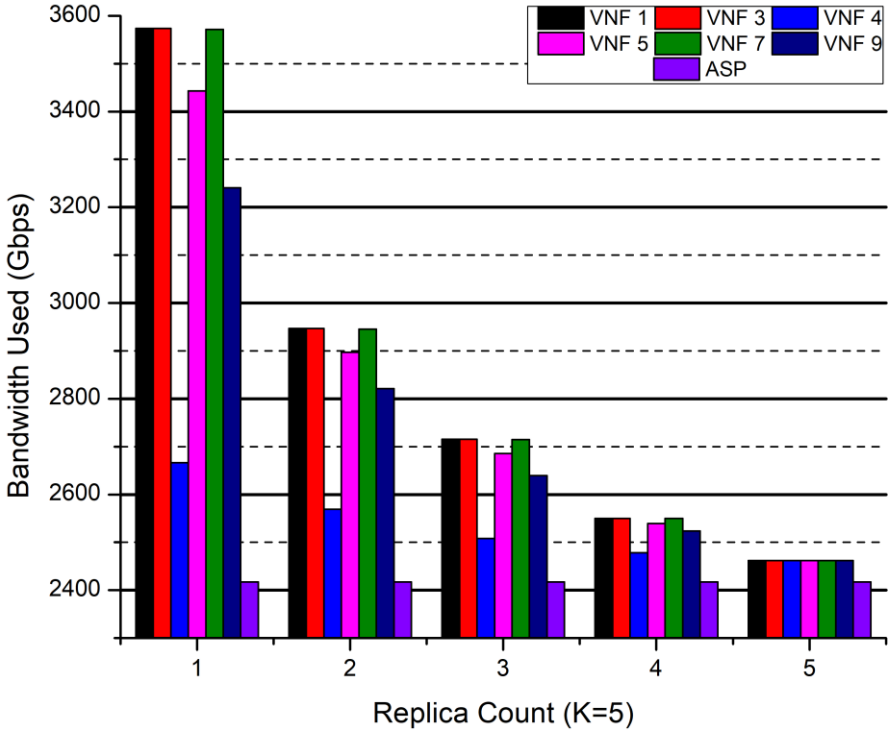


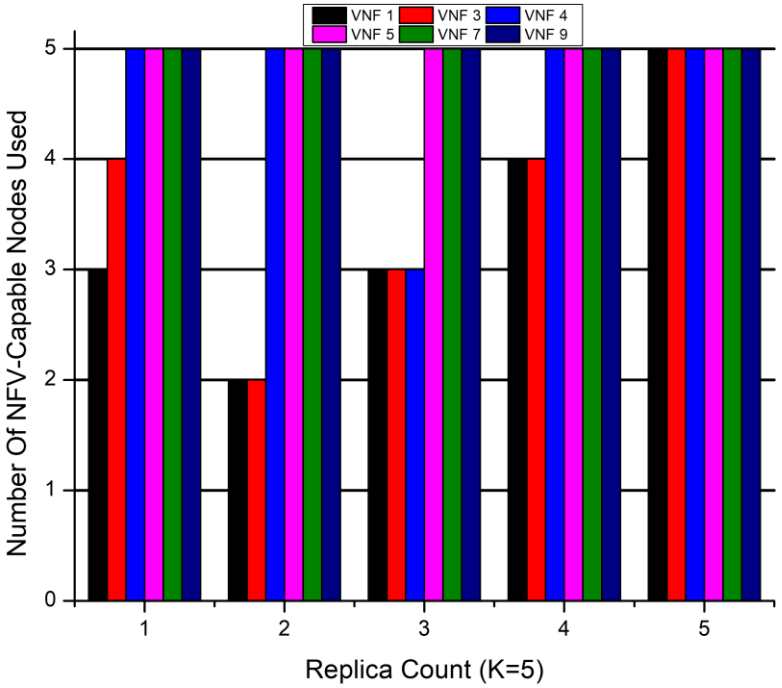


K=5

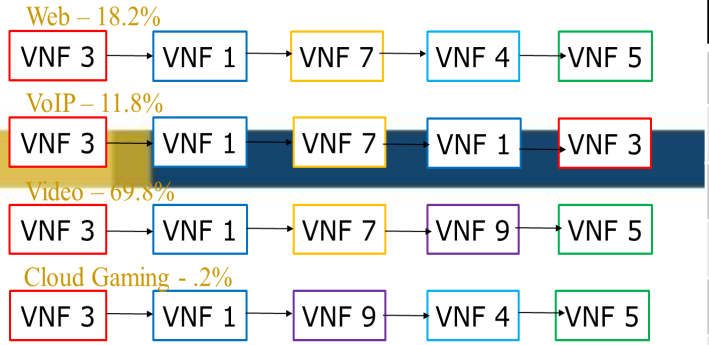


VNF ID	SC's	Traffic Load (Tbps)
3	0,1,2,3	1.118
1	0,1,2,3	1.118
7	0,1,2	.998
5	0,2,3	.882
4	0,3	.184
9	2,3	.700

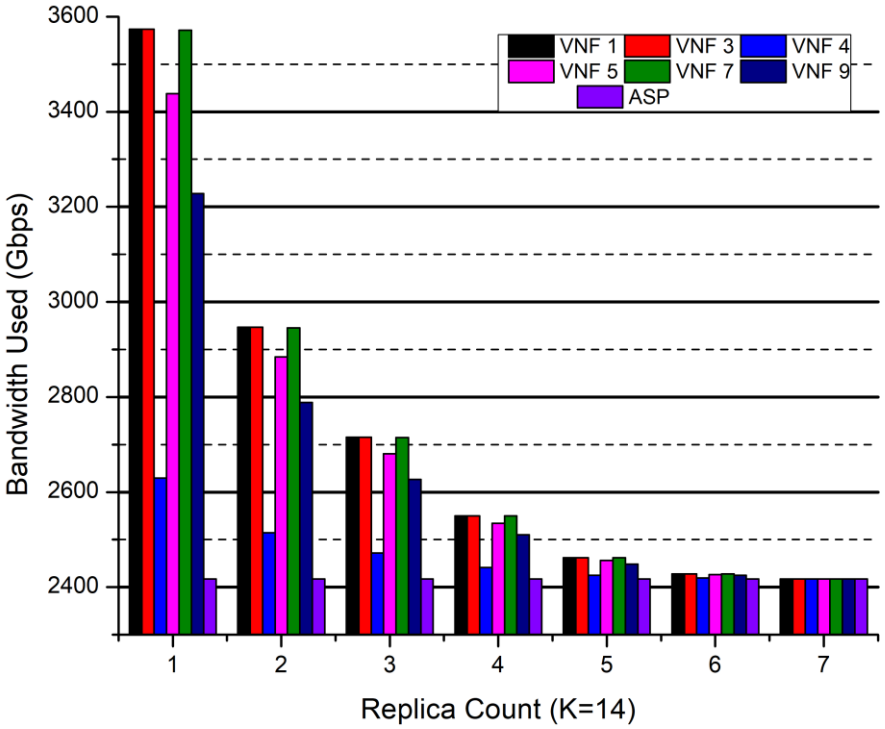


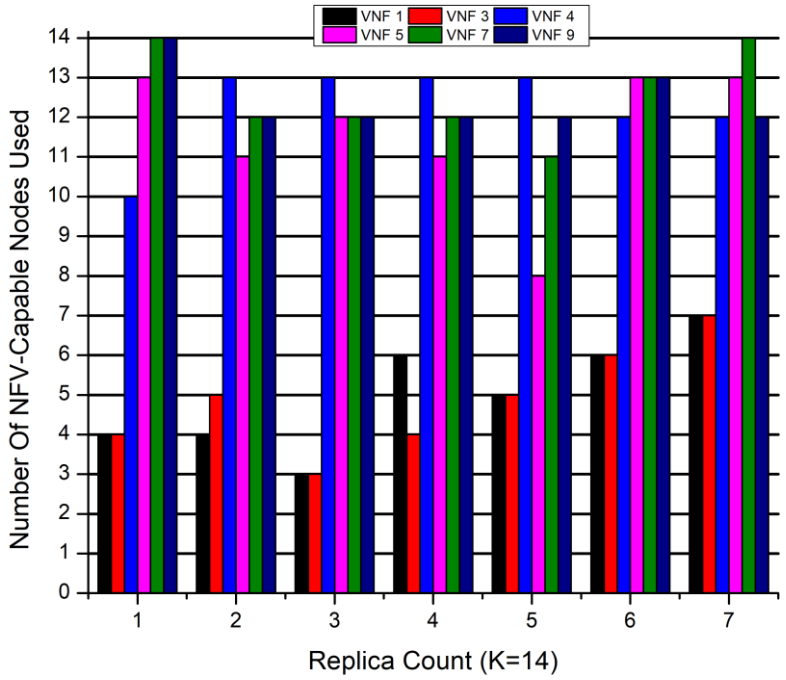


K=14



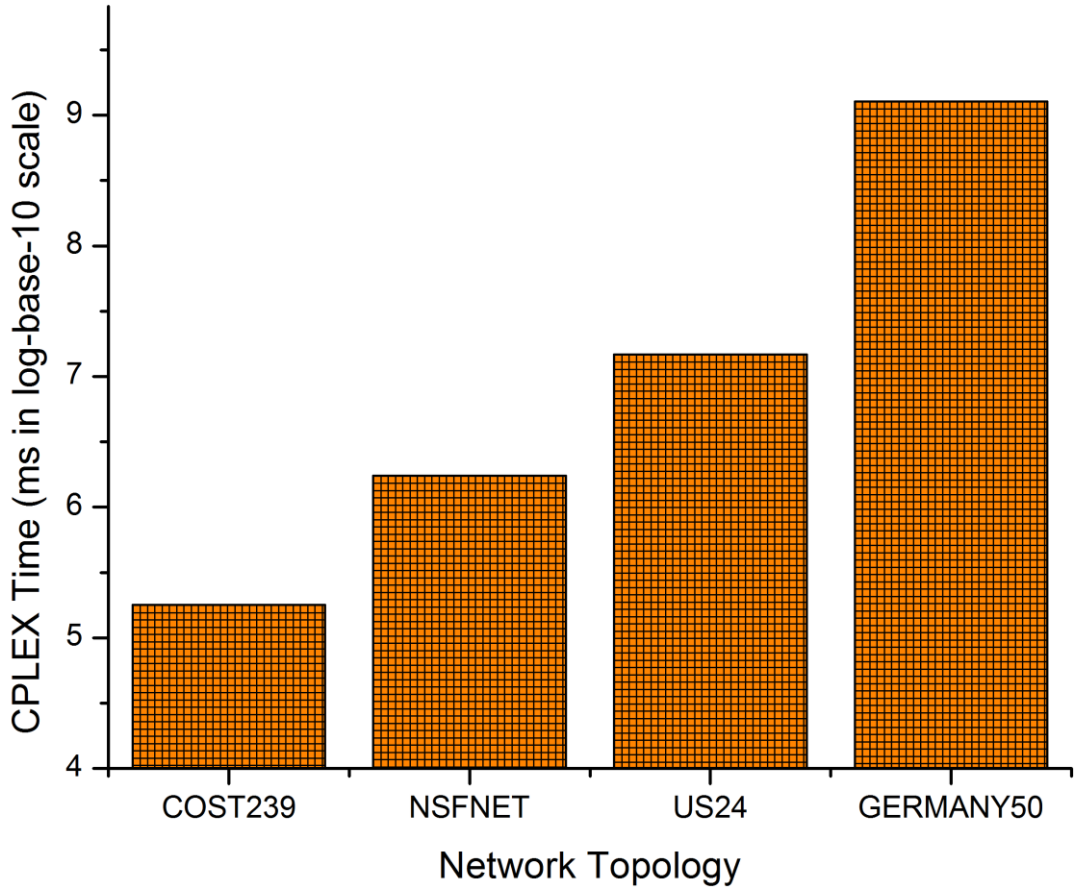
VNF ID	SC's	Traffic Load (Tbps)
3	0,1,2,3	1.118
1	0,1,2,3	1.118
7	0,1,2	.998
5	0,2,3	.882
4	0,3	.184
9	2,3	.700





UCDAVIS
UNIVERSITY OF CALIFORNIA

Scalability of 2 Phase Model



UCDAVIS
UNIVERSITY OF CALIFORNIA

Future Work Directions

- **Cases where distribution of VNFs occur:**
  - Cases where CPU resources are constrained or VNF replicas (because of licenses) are enforced
  - Any additional cases?