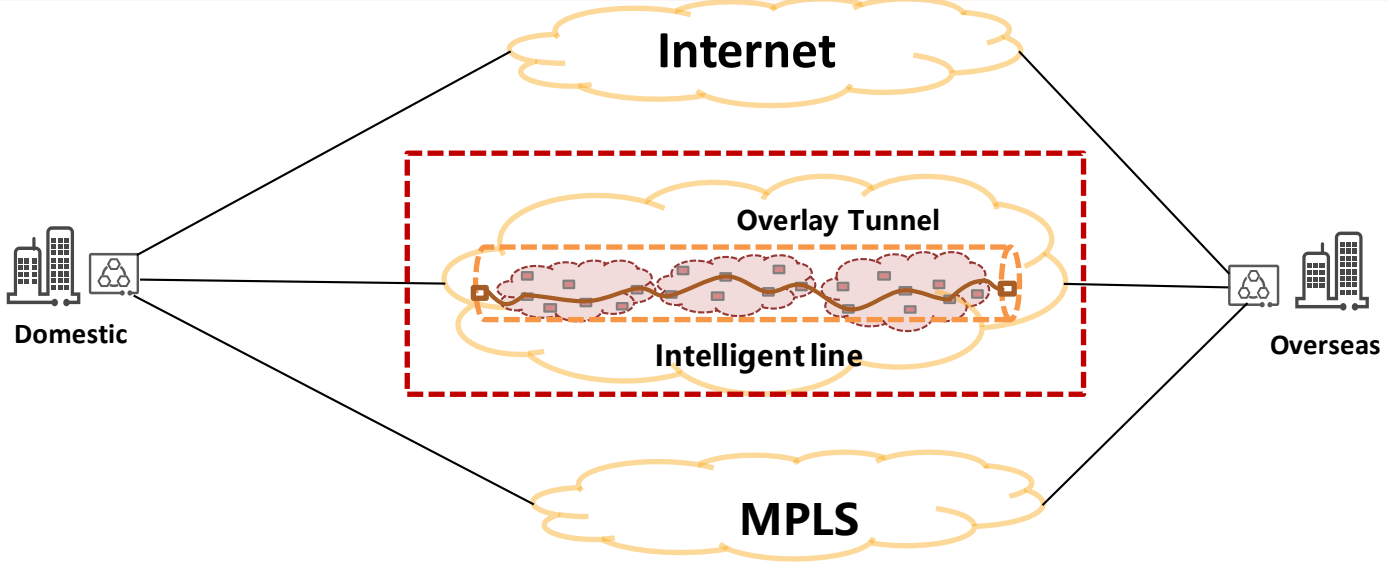
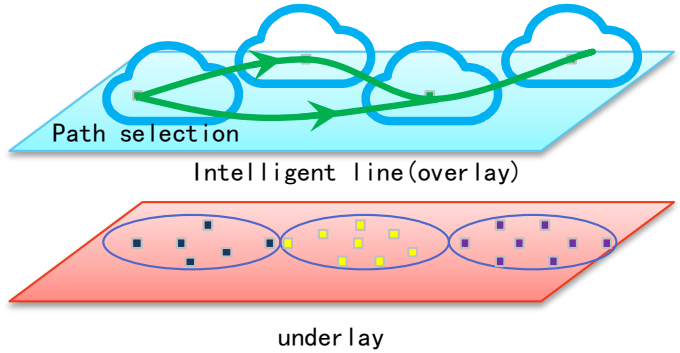
Overlayed Path Segment Forwarding Problem Statement

draft-li-overlayed-path-segment-forwarding

Yizhou Li

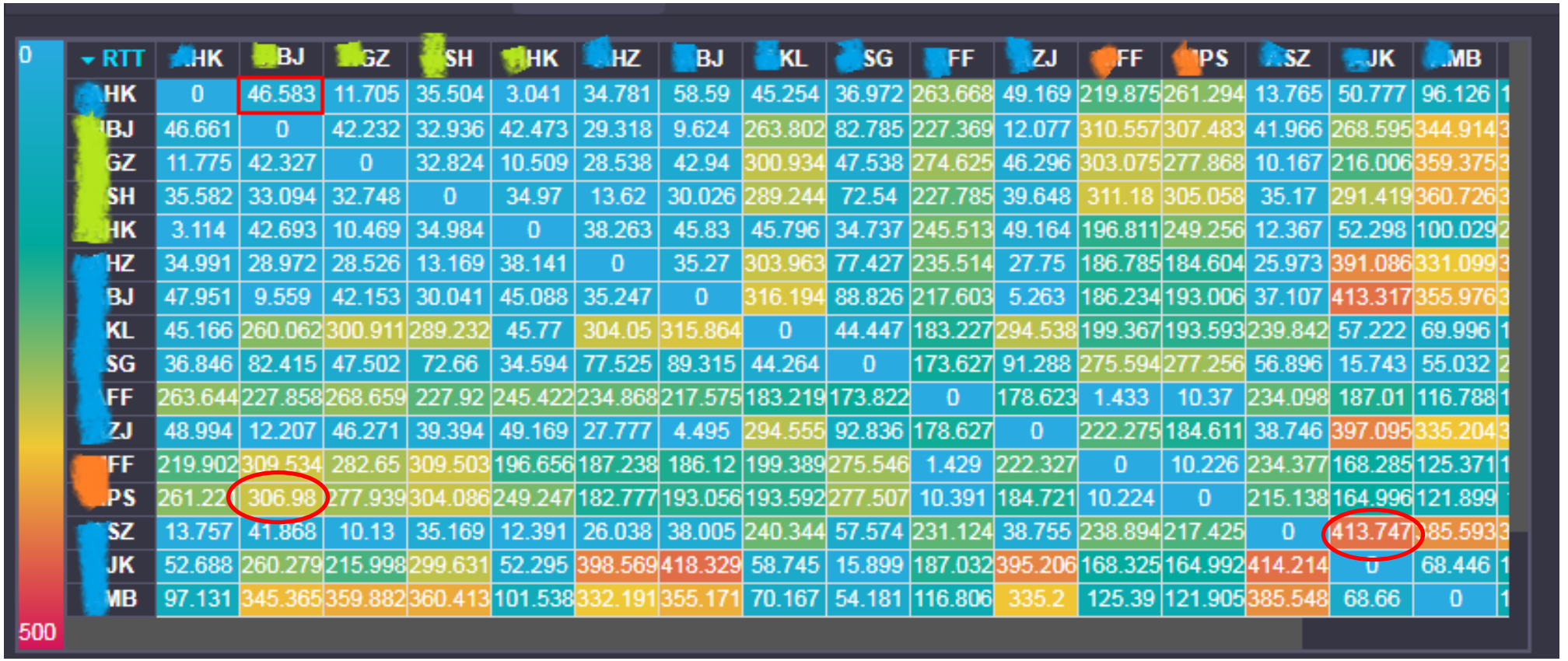
Motivation: Leverage cloud router nodes for best path selection to provide performance closer to leased lines





- Default path does not always give the best latency and throughput
- Now practical: Build a better path via nodes in different geographic sites in the cloud (inexpensive, easy provisioning and scaling, instances with “enhanced network performance” available from cloud provider)
- Experiments: 71% chance of finding a better overlay path based on 37 cloud routers globally

Delays over default path are not always promising

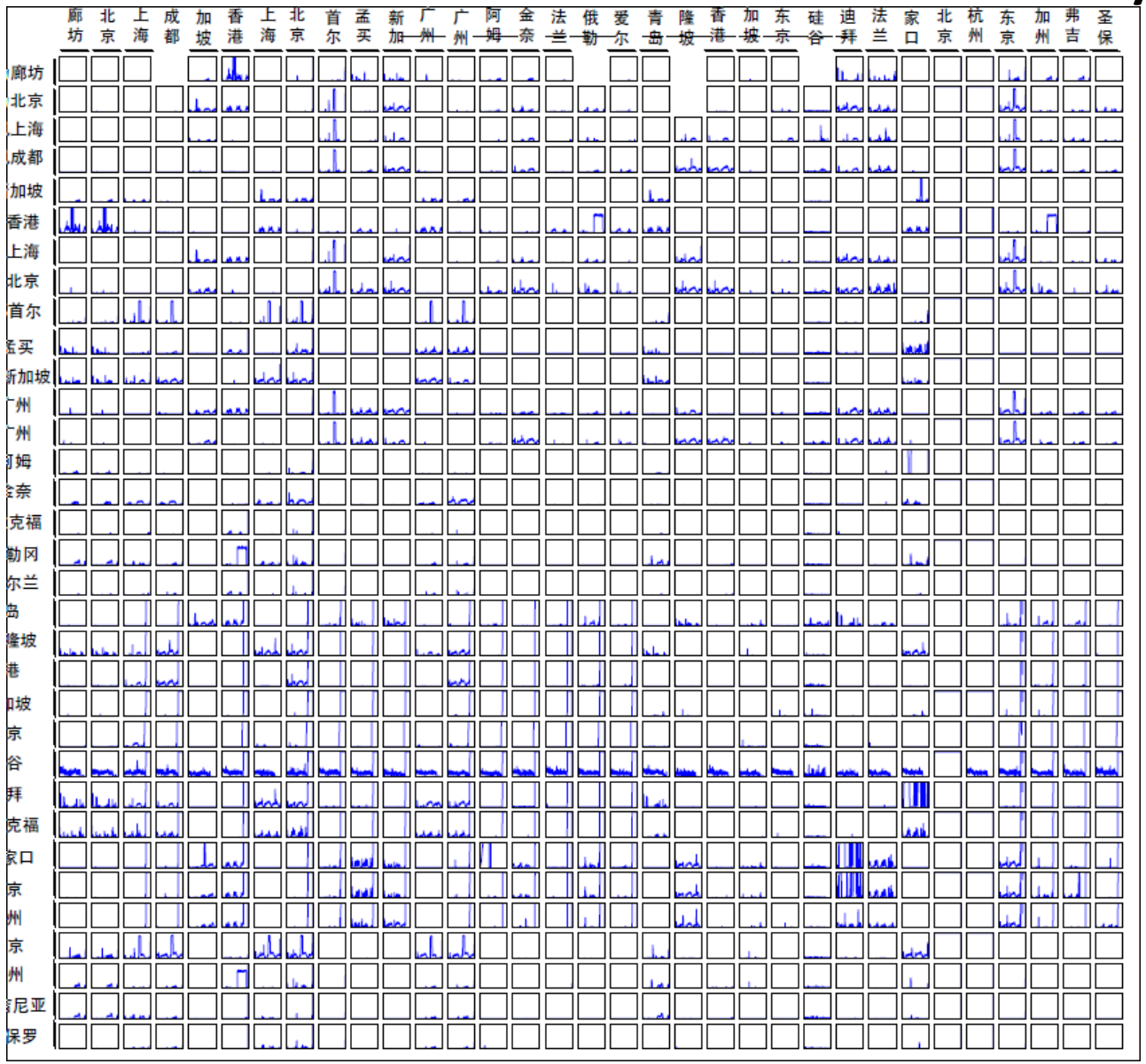


0	RTT	HK	BJ	GZ	SH	HK	HZ	BJ	KL	SG	FF	ZJ	FF	PS	SZ	JK	MB	
	HK	0	46.583	11.705	35.504	3.041	34.781	58.59	45.254	36.972	263.668	49.169	219.875	261.294	13.765	50.777	96.126	1
	BJ	46.661	0	42.232	32.936	42.473	29.318	9.624	263.802	82.785	227.369	12.077	310.557	307.483	41.966	268.595	344.914	3
	GZ	11.775	42.327	0	32.824	10.509	28.538	42.94	300.934	47.538	274.625	46.296	303.075	277.868	10.167	216.006	359.375	3
	SH	35.582	33.094	32.748	0	34.97	13.62	30.026	289.244	72.54	227.785	39.648	311.18	305.058	35.17	291.419	360.726	3
	HK	3.114	42.693	10.469	34.984	0	38.263	45.83	45.796	34.737	245.513	49.164	196.811	249.256	12.367	52.298	100.029	2
	HZ	34.991	28.972	28.526	13.169	38.141	0	35.27	303.963	77.427	235.514	27.75	186.785	184.604	25.973	391.086	331.099	3
	BJ	47.951	9.559	42.153	30.041	45.088	35.247	0	316.194	88.826	217.603	5.263	186.234	193.006	37.107	413.317	355.976	3
	KL	45.166	260.062	300.911	289.232	45.77	304.05	315.864	0	44.447	183.227	294.538	199.367	193.593	239.842	57.222	69.996	1
	SG	36.846	82.415	47.502	72.66	34.594	77.525	89.315	44.264	0	173.627	91.288	275.594	277.256	56.896	15.743	55.032	2
	FF	263.644	227.858	268.659	227.92	245.422	234.868	217.575	183.219	173.822	0	178.623	1.433	10.37	234.098	187.01	116.788	1
	ZJ	48.994	12.207	46.271	39.394	49.169	27.777	4.495	294.555	92.836	178.627	0	222.275	184.611	38.746	397.095	335.204	3
	FF	219.902	309.534	282.65	309.503	196.656	187.238	186.12	199.389	275.546	1.429	222.327	0	10.226	234.377	168.285	125.371	1
	PS	261.22	306.98	277.939	304.086	249.247	182.777	193.056	193.592	277.507	10.391	184.721	10.224	0	215.138	164.996	121.899	
	SZ	13.757	41.868	10.13	35.169	12.391	26.038	38.005	240.344	57.574	231.124	38.755	238.894	217.425	0	413.747	85.593	3
	JK	52.688	260.279	215.998	299.631	52.295	398.569	418.329	58.745	15.899	187.032	395.206	168.325	164.992	414.214	0	68.446	1
	MB	97.131	345.365	359.882	360.413	101.538	332.191	355.171	70.167	54.181	116.806	335.2	125.39	121.905	385.548	68.66	0	1
500																		

- Physical location matters but not always the top factor

* Around 120 virtual nodes.

Loss over default paths between node pairs has different characteristics and vary over time



Certain path has pretty high loss rate all the times

Collected over 3 days.

Problem 1: Slow loss recovery over long haul

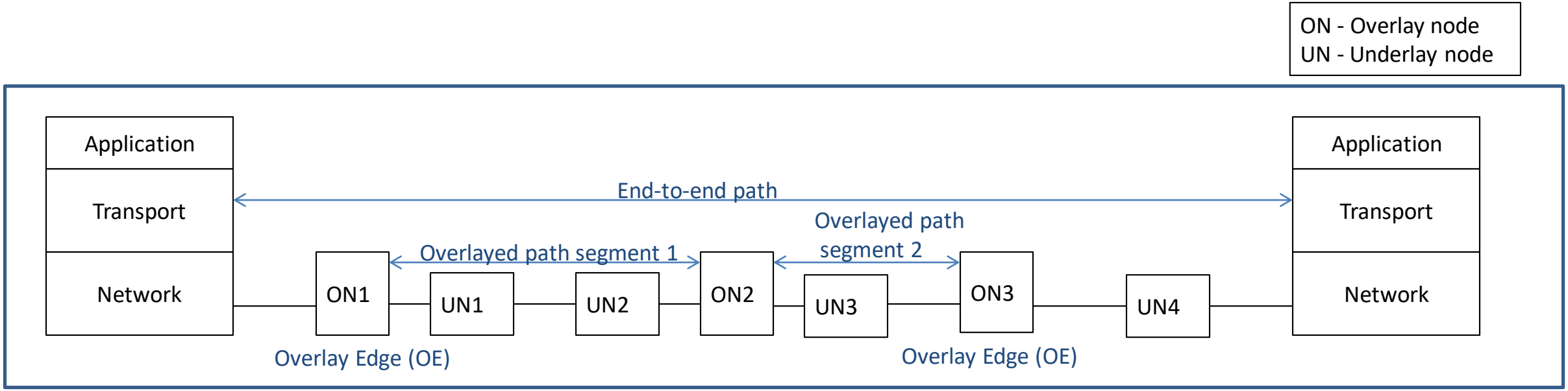
- Intercontinental RTT may take >100 ms. Regional RTT may take tens of ms.
- Interactive app like VoIP, delays perceived at 100ms. Better quality over best-effort Internet is wanted. Some companies exists in the market.
- End-to-end packet recovery is not good enough to meet the req.
- Potential methods:
  - local retransmission over virtual hop
  - FEC over single/multiple path segments
- Why bring it up now?
  - Cloud internet allows better path to be created/selected based on Virtual Nodes (VN).
  - VNs naturally break path to multiple path segments. VN existed or can be easily created and manipulated.
  - Easy to add functions to VN and emerging technologies like virtual IO are making forwarding more efficient

Problem 2: Inaccuracy in sending rate decrease at sender

- Sender decreases its sending rate by a fixed ratio whenever a packet loss is perceived.
- Sender has no much information available to determine how much to decrease otherwise.
- Overlay nodes could provide congestion information to the sender and sender could decide if decreasing sending rate is necessary and how much to decrease.

- Potential methods:
  - Overlay nodes provide richer congestion information to the sender
- Why bring it up now?
  - Good VN features pointed out in last page

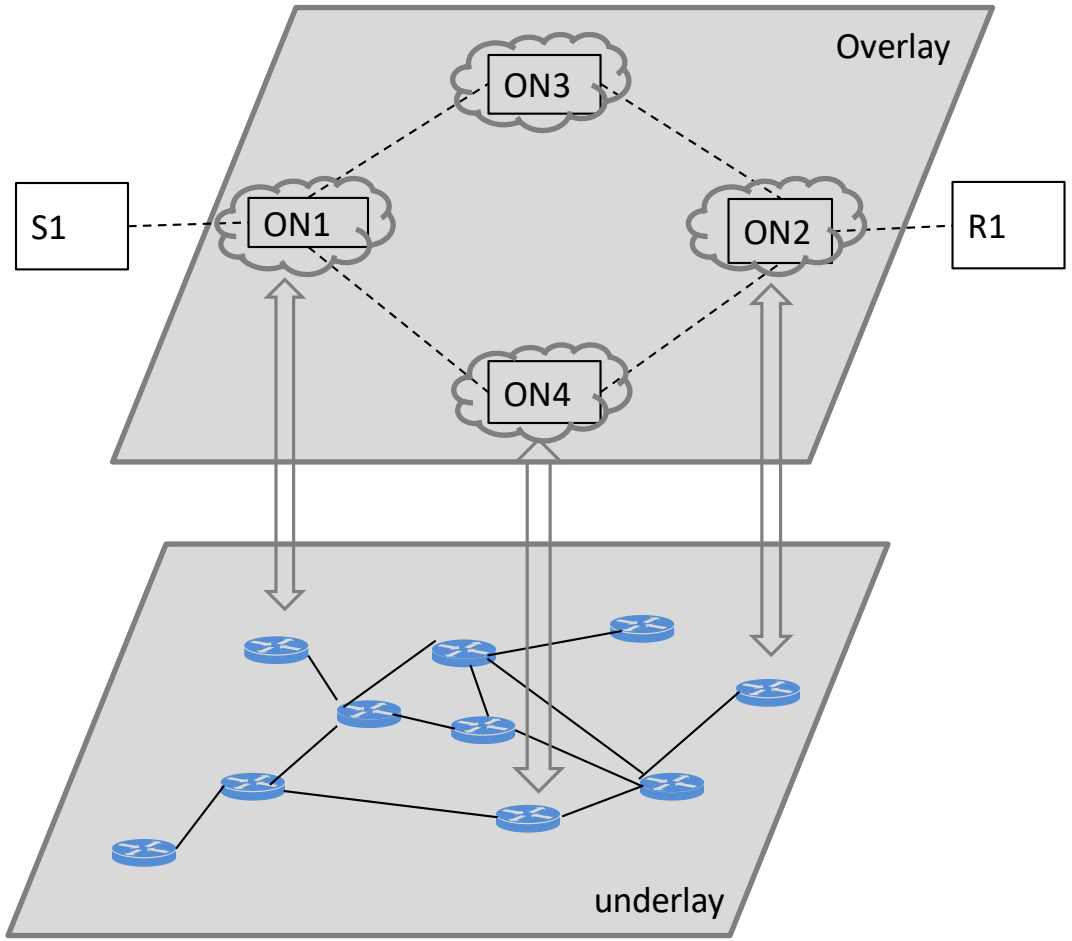
Take this opportunity to do **Localized Optimizations On Path Segment (LOOPS)** for better reliability and throughput



Problems/opportunities:

- Slow recovery over long haul
- Inaccuracy in sending rate decrease at sender
- Impairment/Temporary outage of virtual hop
- Limited capacity of virtual nodes

Elements of a solution



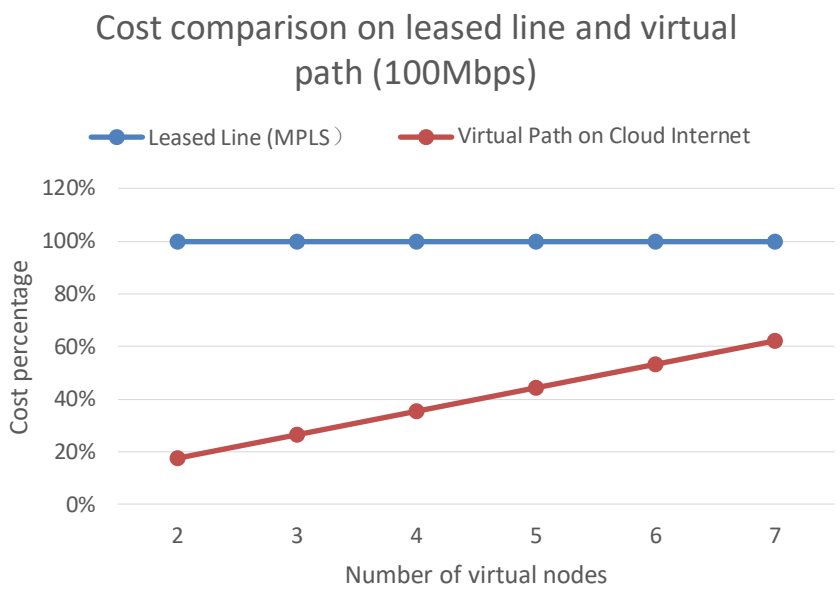
1. Local recovery
   - For entire tunnel (rather than individual flow)
   - Loss detection/indication
   - Measure segment RTT
   - Limited retransmission attempts
   - Control FEC/replication intensity
2. Congestion control interaction
   - Export appropriate CC signaling from LOOPS to e2e transport
   - Support ECN
3. Traffic splitting/recombining
   - For capacity
   - FEC over multiple path segments

Next Steps

- Common elements required for local recovery and traffic splitting.
- Feasible ways to interact with upper layer for CC.
- Overlay protocol extension
  - Individual overlay protocol based; or
  - Generic header
- Comments welcome.

backups

Cost and performance benefit from new tech make multi-segment virtual path feasible



Normally the virtual path takes 2-5 virtual nodes. Cost is fraction of the leased line.

通用网络增强型实例规格族 sn2r

规格族特点

- I/O优化实例
- 仅支持SSD云盘和高效云盘
- 处理器与内存配比为1:4
- 超高网络PPS收发包能力
- 处理器：2.5 GHz主频的Intel Xe
- 实例网络性能与计算规格对应（
- 适用场景：
  - 高网络包收发场景，如视频弹
  - 各种类型和规模的企业级应用
  - 中小型数据库系统、缓存、搜
  - 数据分析和计算
  - 计算集群、依赖内存的数据处

实例规格

实例规格	vCPU

Announcing improved networking performance for Amazon EC2 instances

Posted On: Sep 5, 2017

Enhanced Networking

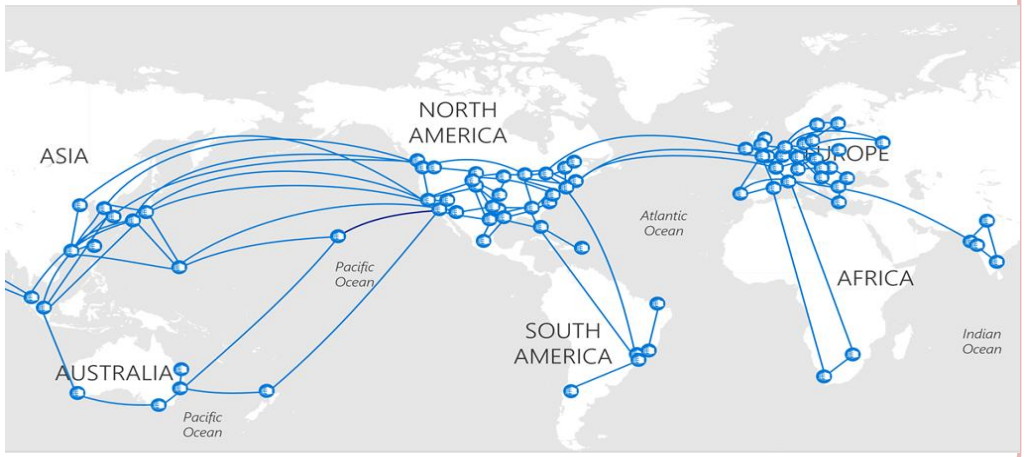
Enhanced Networking enables you to get significantly higher packet per second (PPS) performance, lower network jitter and lower latencies. This feature uses a new network virtualization stack that provides higher I/O performance and lower CPU utilization compared to traditional implementations. In order to take advantage of Enhanced Networking, you should launch an HVM AMI in VPC, and install the appropriate driver. For instructions on how to enable Enhanced Networking on EC2 instances, see the Enhanced Networking on Linux and Enhanced Networking on Windows tutorials. For availability of this feature by instance, or to learn more, visit the Enhanced Networking FAQ section.

Instances with enhanced networking feature to provide better network forwarding performance becoming available from cloud providers.

Measurement based overlay path selection

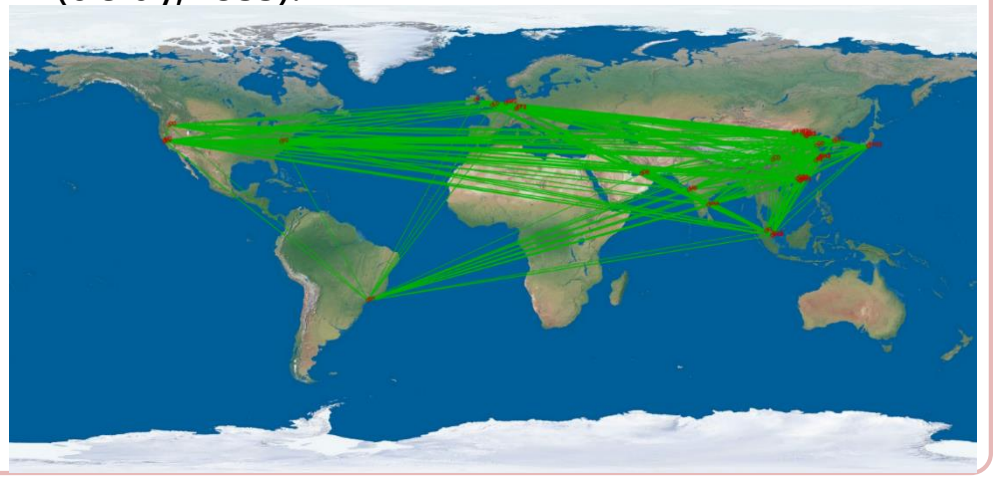
VN creation on Cloud Internet

- Virtual nodes located with geographic location diversity (100-200 sites)

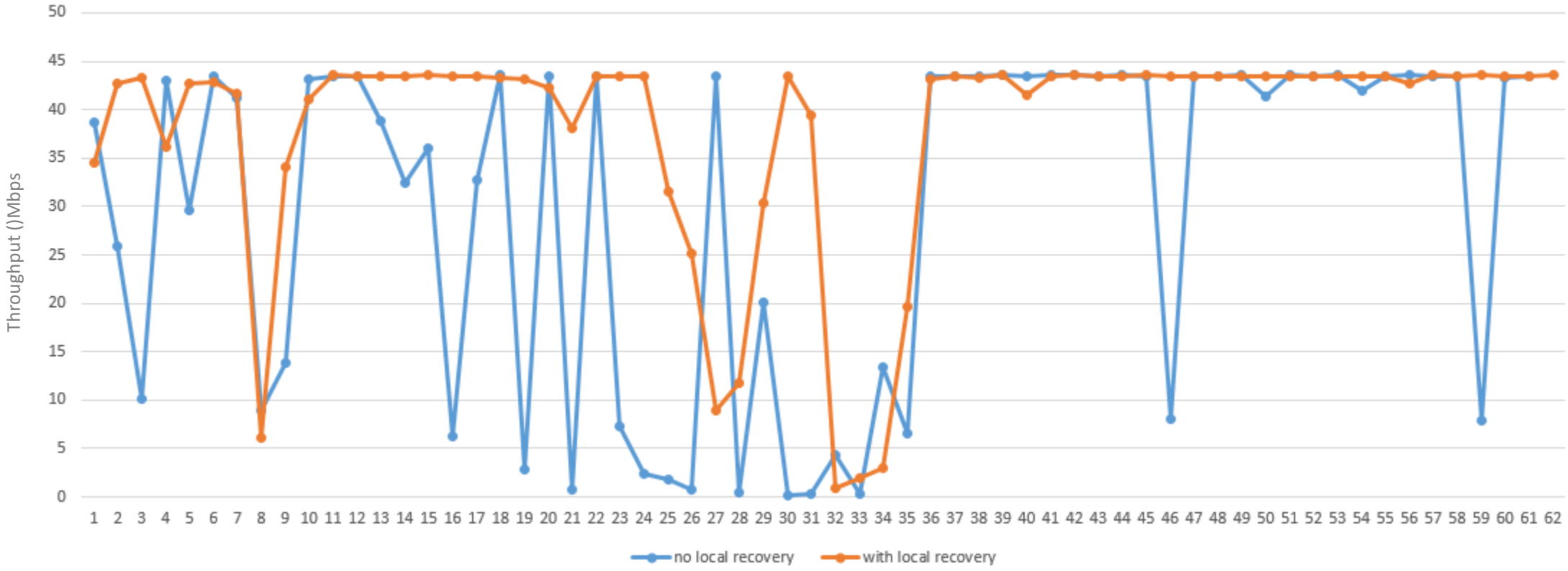


Measurement based Node Selection

- Static measurement: measure between node pairs as the base
- Real time measurement (in-band and out-of-band): select the VNs as intermediate nodes to form the candidate best paths (delay, loss).



Preliminary test of local retransmission with overlay path



Local Retransmission generally gives better throughput

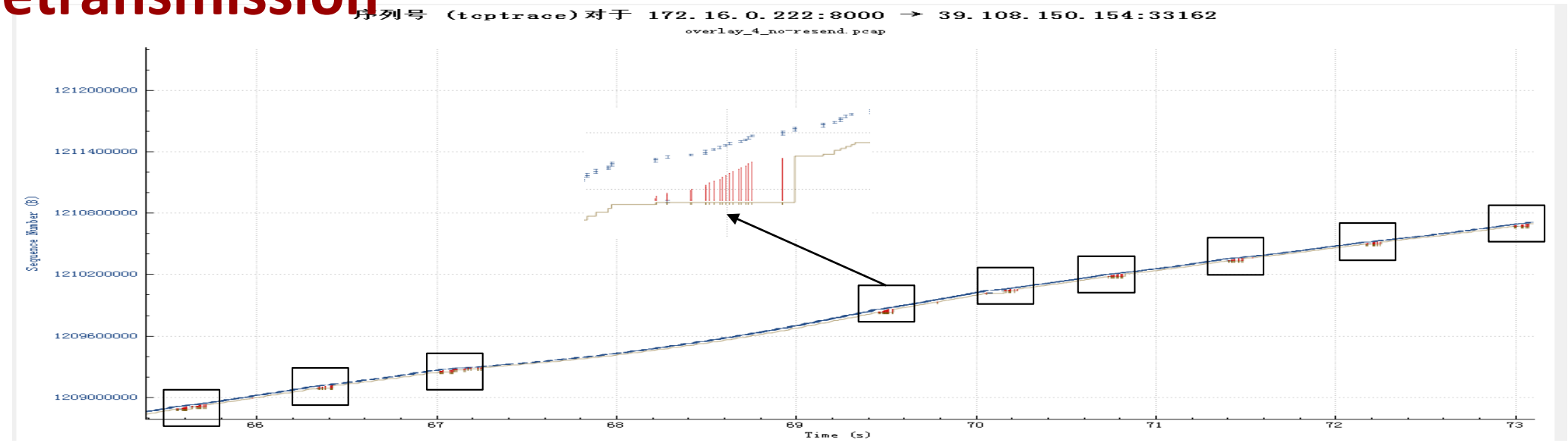
Path: Jakarta to Shenzhen, 4 ONs

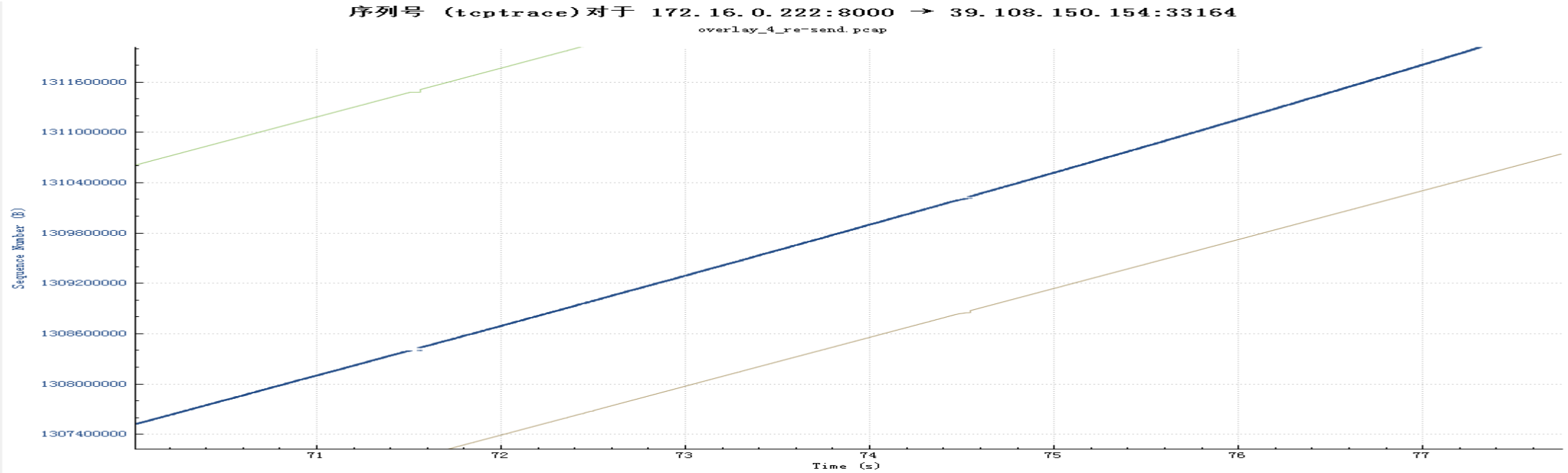
X-Axis: test cycles

FTP file downloading

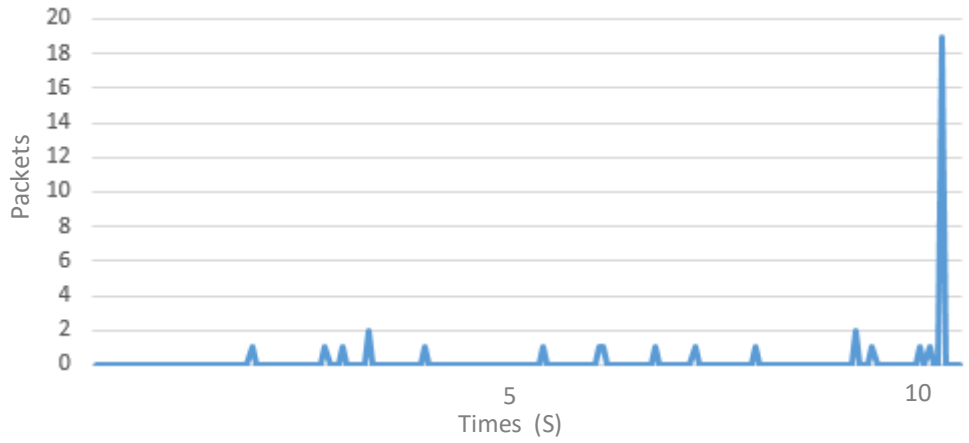
Back-to-back for 120 seconds each testing, then wait for 5 minutes

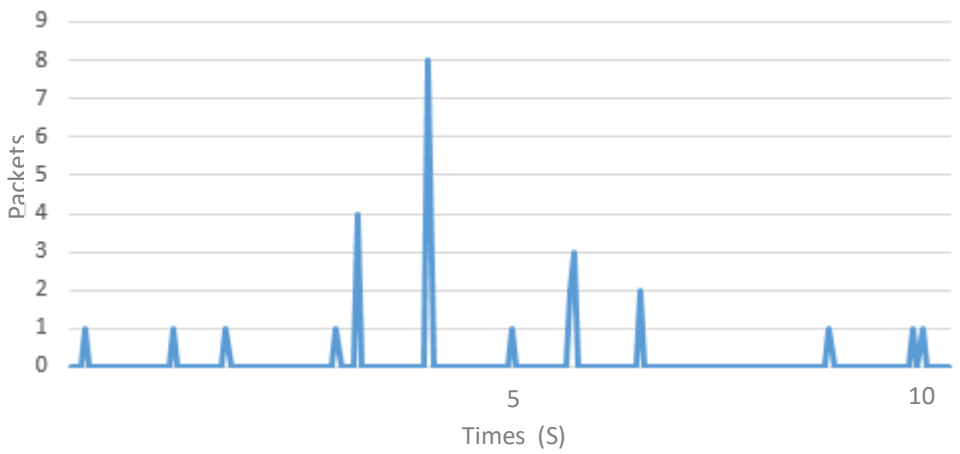
Loss detected by sender is much higher in case of no local retransmission

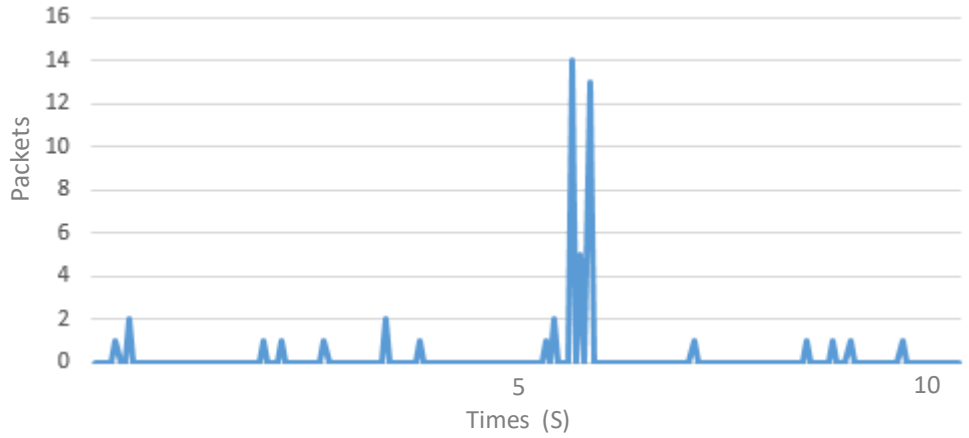




Loss observed on virtual nodes lasts for very short time







Non-persistent loss is common.

*50ms sampling