Failure Resiliency of NFV Services

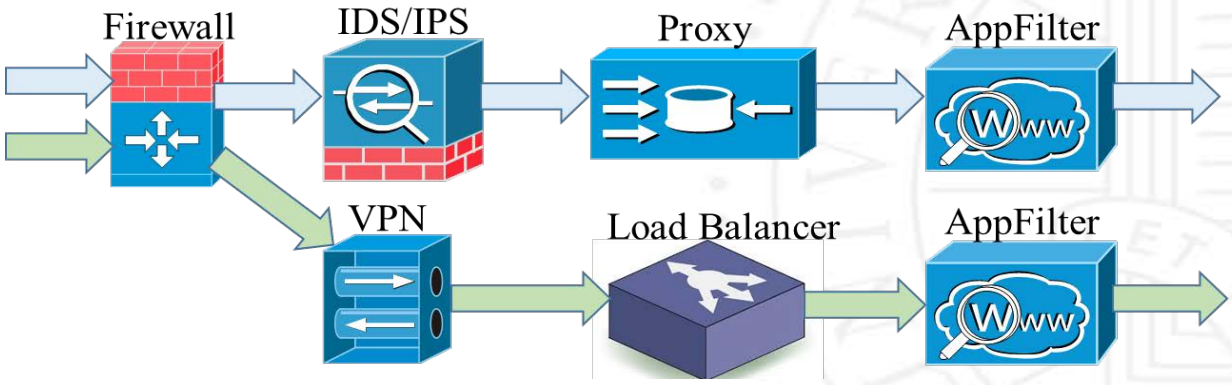
K. K. Ramakrishnan
University of California, Riverside, CA
In collaboration with
Sameer Kulkarni and Tim Wood (GWU)

UCR
UNIVERSITY OF CALIFORNIA, RIVERSIDE

Background: Network Functions (NFs)

- Middleboxes are everywhere in communication service provider (CSP) networks.
  - Software based NFs are fast replacing the purpose-built hardware middleboxes.
  - Run these software based network functions (NFs) on standardized server machines.

- Network function chains:
  - Provide different network services; implement variety of network policies.



NF Service Availability is Important

- NFs: are in a flow's path, like a bump-in-the-wire - affect the service.
- Network service outages incur significant loss of revenue [1]!
  - Recent survey on 200 companies reported loss of > $26.5billion/yr. due to network outages (nearly $8K per minute).

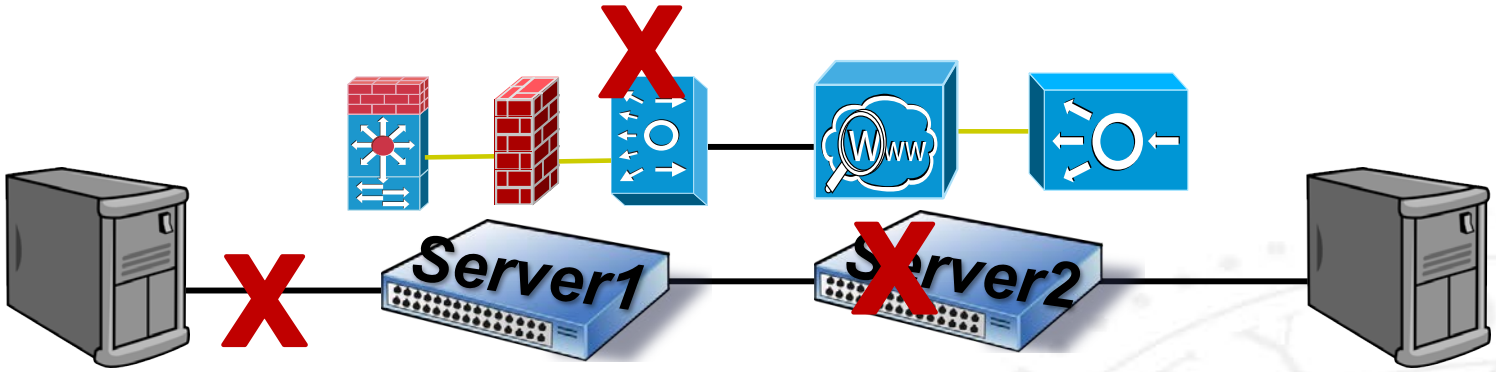
Need to build Failure Resilient NF Services!

- Communication Service Providers demand >= 99.999% availability[2]!

	Middleboxes		Standard Server Machines [5]
	CG-NAT [3]	Firewall [4]	
Availability(%)	99.999	99.999	< 99.9

[1] https://www.linktek.com/cost-of-network-downtime/
[2] https://nae.global/en/network-architecture-and-spectrum-innovations-for-5g/
[3,4] http://www.f5.com/pdf/products/big-ip-cgnat-datasheet.pdf, big-ip-advanced-firewall-manager-datasheet.pdf
[5] https://www.ctl.io/legal/centurylink-cloud/sla/, https://aws.amazon.com/s3/sla/ compute/sla/

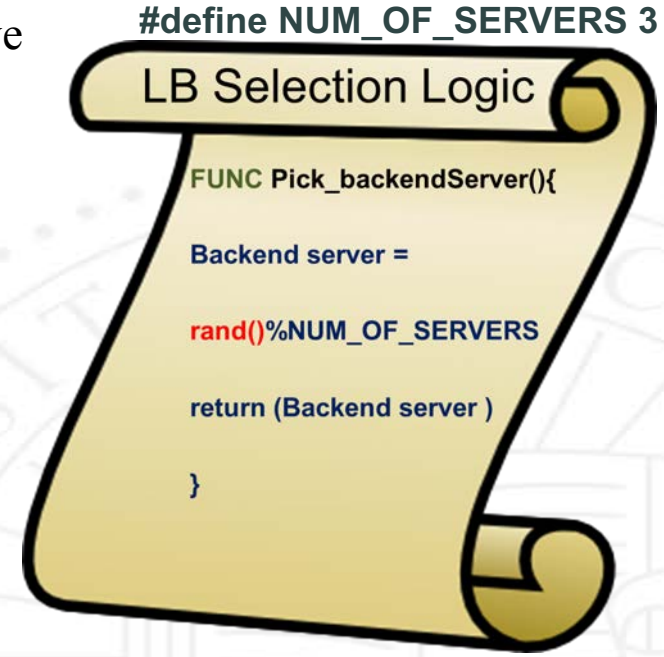
Problem Statement: NF Failure Resiliency



- Failure Resiliency framework for NF and NF chains:
  - Must address different kinds of failures:
    - Software (NF Instance) failures
    - Hardware (Link, Node) failures.
- ***How to quickly detect these different kinds of failures?***
- ***How to provide efficient and correct chain-wide failover to redundant (replica, secondary) service instances?***

Correctness Challenge: Non-determinism

- NFs often exhibit Non-Determinism [FTMB, SIGCOMM'15]
  - Given two identical middleboxes, with identical input, their respective output (processing result) can differ.
- Sources of non-determinism:
  - Software:
    - Shared variables: ordering of threads.
    - System dependent APIs
  - Hardware:
    - Clock access, resource contention.
  - Probabilistic decisions on packets
    - RED/REM.



Must address Non-determinism!

[FTMB, SIGCOMM'15] Sherry J, et.al, "FTMB: Fault Tolerant Middleboxes, SIGCOMM 2015.

Requirements: NFV Resiliency Solution

- Correctness:
  - Loss-free state updates to replica.
- Low Overhead:
  - Minimal impact on performance (Latency and Throughput) for Normal (failure free) operation.
- Quick Recovery:
  - Interruption free failover – minimal service impact
- Generality:
  - Work for different NFs and deployment models.

REINFOR

Efficient Resiliency framework for NFV chains.

- Efficient and Correct state migration: *External Synchrony*[1]
  - Separation of ***NF** state* and ***chain-wide processing progress*** state.
  - Separation of ***Deterministic*** and ***Non-deterministic*** packet processing.
  - Non-blocking pipelined chain wide processing with batch commits.
- Fast Failure Detection (NF instances, Link and Node):
  - Link and Node Failures: Extend ***BFD***[2] with active traffic piggybacking.
  - Local NF Instance failure: Lightweight probe based monitoring.
- Redundancy Approach:
  - *Primary-Hot standby* (local NF instance; remote node standby for chain-wide redundancy).
  - *Distinct failover schemes* for (local NF instance, link &node) failures.

[1] Nightingale, Edmund B., et al. "Rethink the sync." OSDI 2006, *ACM ToCS 2008*.
[2] BKatZ D, & Ward D, Bidirectional Failure Detection, RFC5880, RFC5881.

REINFORCE

Efficient Resiliency framework for NFV chains.

- Fast and correct failover:
  - Chain-wide remote node failover in < ***5 ms***.
  - Local NF instance failover in ***< 100 µs***
- Low overhead on failure free operations:
  - ***< 15%*** performance impact on Normal (Failure Free) operation.
- Fast failure detection (NF Instances, Link & Node):
  - ***~50 µs*** for detecting NF failures
  - ***< 3 ms*** for detecting Link/Node failures.